

ESSAYS ON THE PREDICTABILITY AND VOLATILITY
OF ASSET RETURNS

A Dissertation

by

STEFAN A. JACEWITZ

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2009

Major Subject: Economics

ESSAYS ON THE PREDICTABILITY AND VOLATILITY
OF ASSET RETURNS

A Dissertation

by

STEFAN A. JACEWITZ

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Joon Y. Park
Committee Members,	Yoosoon Chang
	Hwagyun Kim
	Michael Gallmeyer
Head of Department,	Larry Oliver

August 2009

Major Subject: Economics

ABSTRACT

Essays on the Predictability and Volatility
of Asset Returns. (August 2009)

Stefan A. Jacewitz, B.A., University of Oklahoma

Chair of Advisory Committee: Joon Y. Park

This dissertation collects two papers regarding the econometric and economic theory and testing of the predictability of asset returns. It is widely accepted that stock returns are not only predictable but highly so. This belief is due to an abundance of existing empirical literature finding often overwhelming evidence in favor of predictability. The common regressors used to test predictability (e.g., the dividend-price ratio for stock returns) are very persistent and their innovations are highly correlated with returns. Persistence when combined with a correlation between innovations in the regressor and asset returns can cause substantial over-rejection of a true null hypothesis. This result is both well documented and well known. On the other hand, stochastic volatility is both broadly accepted as a part of return time series and largely ignored by the existing econometric literature on the predictability of returns. The severe effect that stochastic volatility can have on standard tests are demonstrated here. These deleterious effects render standard tests invalid. However, this problem can be easily corrected using a simple change of chronometer. When a return time series is read in the usual way, at regular intervals of time (e.g., daily observations), then the distribution of returns is highly non-normal and displays marked time heterogeneity. If the return time series is, instead, read according to a clock based on regular intervals of volatility, then returns will be independent and identically normally distributed. This powerful result is utilized in a unique way in each chapter of this dissertation. This time-deformation technique is combined with the Cauchy t -test

and the newly introduced martingale estimation technique. This dissertation finds no evidence of predictability in stock returns. Moreover, using martingale estimation, the cause of the Forward Premium Anomaly may be more easily discerned.

To Keli and Maksym; my best friends.

ACKNOWLEDGMENTS

I would like to acknowledge the financial support of Texas A&M University, the emotional support of my wife, and the academic support of my advisor.

TABLE OF CONTENTS

CHAPTER		Page
I	INTRODUCTION	1
II	A TEST OF STOCK RETURN PREDICTABILITY USING THE CAUCHY T -RATIO IN VOLATILITY TIME	2
	A. Introduction	2
	B. Background and Main Issues	6
	1. Persistence and Endogeneity of Predictors	7
	2. Nonstationary Stochastic Volatility in Returns	10
	3. Other Issues	12
	C. A Novel Approach	14
	1. Efficient Tests	20
	2. A Robust Test - The Cauchy Estimator	23
	D. A New Test and Its Statistical Theory	30
	1. A New Test	30
	2. Asymptotic Theory	34
	E. Monte Carlo Simulations	37
	F. Empirical Results	46
	G. Conclusion	54
III	TESTING FOR NO ARBITRAGE IN CONTINUOUS TIME: A RESOLUTION TO THE FORWARD PREMIUM ANOMALY	57
	A. Introduction	58
	1. Theory	61
	2. The Anomaly	64
	B. A Continuous Time Model of the Foreign Currency Market and Forward Exchange Rates	67
	C. Econometric Methodology	75
	1. Econometric Model	75
	2. Time Change	77
	3. Martingale Estimation	82

CHAPTER	Page
D. Empirical Results	86
1. Data	87
2. Estimation Procedure	88
3. Estimates	90
4. Identification	96
5. Modeling the Market Price of Risk	96
E. Conclusion	103
IV CONCLUSION	104
REFERENCES	105
APPENDIX A	113
APPENDIX B	117
APPENDIX C	120
APPENDIX D	122
VITA	123

LIST OF TABLES

TABLE		Page
I	Estimates based on Campbell and Yogo's (2006) efficient test of predictability performed on the time-changed data.	22
II	Estimates for Chen and Deo's (2008) Restricted Maximum Likelihood (REML) test performed on the time-changed data.	22
III	Sizes in calendar time. This table demonstrates the effects of a structural break in volatility on the OLS and Cauchy t -tests, the Bonferroni Q -test, and the REML test. The nominal level is 5% and all values are given in terms of percentages. The experiment is replicated 10,000 times.	43
IV	Sizes in volatility time. This table demonstrates the detrimental effects of a structural break in volatility on the OLS and Cauchy t -tests, the Bonferroni Q -test, and the REML test when applied to the volatility time data. The nominal level is 5% and all values are given in terms of percentages. The experiment is replicated 10,000 times.	44
V	Sizes in calendar time. This table demonstrates the effects of nonstationary volatility on the OLS and Cauchy t -tests, the Bonferroni Q -test, and the REML test. The nominal level is 5% and all values are given in terms of percentages. The experiment is replicated 10,000 times.	45
VI	Sizes in volatility time. This table demonstrates the effects of nonstationary volatility on the OLS and Cauchy t -tests, the Bonferroni Q -test, and the REML test when applied to the time changed data. The nominal level is 5% and all values are given in terms of percentages. The experiment is replicated 10,000 times.	47
VII	Results of selecting parameter values which achieve the highest degree of normality. "CvM" refers to the Cramér-von Mises distance from a standard normal distribution.	52

TABLE	Page
VIII	Main empirical results for return predictability. D/P and E/P represents the dividend- and earnings-price ratios, respectively. TC indicates a time change was employed. 53
IX	Subsample empirical results for return predictability. 53
X	Jump robustness check. This table provides estimates of the predictability parameter when intervals which have test statistics below the 10% critical value are removed. 55
XI	Model 0 with Δ chosen by minimizing the CvM statistic. 91
XII	Model 1 with Δ chosen by minimizing the CvM statistic. 92
XIII	Model 2 with Δ chosen by minimizing the CvM statistic using the standard deviation of the full sample of the risky return as the measurement of risk. 93
XIV	Model 2 with Δ chosen by minimizing the CvM statistic and with ν_t estimated using a rolling standard deviation for the previous month. 94
XV	Model 3 with Δ selected by minimizing the CvM statistic and estimating risk via rolling standard deviation over the previous year. 95
XVI	Model 2 with Δ chosen by minimizing the CvM statistic and with ν_t estimated using the nonparametric technique from Stanton (1997). 100
XVII	This table compares results based on the standard OLS regression to the results from Cauchy estimation with a time change when there is a structural break in volatility. Here, “TC-Cauchy” represents the Cauchy estimate of the time-changed data and $T = 25$. All values are given in terms of percentages. 118
XVIII	This table compares results based on the standard OLS regression to the results from Cauchy estimation with a time change. Here, “TC-Cauchy” represents the Cauchy estimate of the time-changed data and $T = 25$. All values are given in terms of percentages. 119

TABLE		Page
XIX	This table provides the selected values of N and equivalently Δ for each country based on minimizing the one dimensional CvM statistics of parameter estimates. The median estimates of β across the range of Δ is also reported.	121

LIST OF FIGURES

FIGURE		Page
1	Calender versus volatility time. This figure demonstrates the conversion between calender time and volatility time. Instead of proceeding by one interval of time (i.e., one month), we proceed by one interval of volatility between observations. For illustrative purposes, not all of the random time observations are shown here. . .	17
2	Power comparison. A comparison of local asymptotic power for the Cauchy t -test, the infeasible t -test, and the Bonferroni Q -test. Here, let $b = T\beta$ to account for the sample size. This figure is generated using 20,000 replications of a 500 observation sample. . . .	29
3	Data plot. This figure plots the data, with monthly, quarterly, and yearly data going from left to right. The regressors are displayed on the top row and the stock returns are displayed on the bottom. For the regressors, the solid blue line represents the dividend-price ratio, while the dotted green line represents the earnings price ratio. .	48
4	Data plot and kernel estimates of return distributions before and after the time change. The dashed line represents a standard normal distribution and the blue line gives the estimated distributions for innovations normalized by their respective standard deviations.	49
5	Kernel estimates of innovation distributions for the parameter values which achieve the most normal innovations in the regressor under standard time sampling and under random time sampling. These are distributions for the dividend-price ratio. For the earnings-price ratio, the results are entirely similar. The dashed line represents a standard normal distribution and the blue line gives the estimated distributions for innovations normalized by their respective standard deviations.	51

FIGURE		Page
6	Kernel density estimates of the return distributions before and after the time change has been performed. The top six panels give the density estimates of the standard time returns for each country. The bottom six panels give the density estimates for the random time returns. The dashed line represents a standard normal distribution.	78
7	Daily rolling standard deviation of returns for the previous year. . . .	79
8	Calender versus volatility time. This figure demonstrates a conversion from calender time to volatility time for the British pound. The vertical axis maps the realized volatility in regular increments. Meanwhile, on the horizontal axis is the time between observations is a random variable.	83
9	The Cramér-von Mises distance for various values of the parameters β and γ holding $\alpha = 0$	97
10	A contour plot of the Cramér-von Mises distance for various values of the parameters β and γ holding $\alpha = 0$	98
11	The Cramér-von Mises distance for various values of the parameters β and γ holding $\alpha = 0$	101
12	A contour plot of the Cramér-von Mises distance for various values of the parameters β and γ holding $\alpha = 0$	102

CHAPTER I

INTRODUCTION

Methodologically, the uniting feature of this dissertation is time deformation via a clock based on volatility time rather than calendar time. Returns in both the stock and currency markets are widely accepted as displaying stochastic volatility. The presence of stochastic volatility can have devastating effects on standard statistical analysis. Observing the data using a volatility clock complete eliminates any stochastic volatility present in the data and leaves returns which are very well behaved. Empirically, this dissertation focuses on predictability of asset returns. The predictability of stock returns is perhaps the most important research topic in Finance. Likewise, the ability of forward premium to correctly predicts exchange rate returns is firmly established as one of the most important questions in International Finance.

The first chapter concerns itself with testing whether stock returns may be predicted using a few standard ratios commonly presumed to have strong predictive power. Persistence and endogeneity can cause over-rejection of a true null. At the same time, the existence of stochastic volatility, a condition entirely ignored by the current literature, can aggravate the distortionary effect. The standard statistical issues are easily addressed by the Cauchy t -ratio, while stochastic volatility is addressed using a change of chronometer.

The second chapter deals instead with whether the forward premium predicts future currency returns. Similar statistical issues exist in this topic as well. Instead of the Cauchy t -statistic, the new Martingale Estimator is used to test the well known Forward Premium Anomaly.

This dissertation follows the style of *Econometrica*.

CHAPTER II

A TEST OF STOCK RETURN PREDICTABILITY USING THE CAUCHY T -RATIO IN VOLATILITY TIME

This chapter proposes a novel approach to testing for predictability in stock returns. When the new technique, a change to volatility time and a Cauchy t -ratio, is applied to the data, we find no evidence for stock return predictability. This is a quite surprising result since a vast amount of previous literature finds that stock returns are not only predictable, but highly so. In this paper, it is shown that certain ubiquitous data characteristics, such as the largely unaddressed presence of stochastic volatility in stock returns, as well as the well known persistence in the commonly used predictors and a strong correlation between innovations in the regressor and stock returns, have a substantial impact on traditional hypothesis testing. Our new technique for testing predictability is uniquely suited to each of these widely recognized characteristics of predictive regression data. The technique consists of a simple time change to volatility time to accommodate a general form of stochastic volatility in stock returns, and instrumental variable estimation to allow for a wide range of endogenous nonstationary covariates.

A. Introduction

Are stock returns predictable? Both sides of this contentious debate have been prolifically active for decades. In this paper, we find absolutely no evidence supporting stock return predictability. This result is surprising since much of the research in this area has consistently concluded that stock returns are not only predictable, but highly so. Indeed, Cochrane (2005) dubs this well known result as one of the “new facts in finance”. This conclusion has lead to a fruitful and still expanding literature

on two fronts. Many theorists have developed models supporting such predictability. On the other hand, many econometricians have been troubled by the techniques used to test for predictability. In this paper, this paper takes the latter approach demonstrating some of the weaknesses in the most popular tests and offering an original and stronger alternative. In the past, the literature has produced dozens of papers most of which rely on the same basic machinery to address predictability. This paper offer a completely new approach. We combine a simple nonlinear instrument and a volatility clock. This test is uniquely suited to address the problematic characteristics found in predictive regression data. Unlike nearly all previous work, however, we are left with no evidence for predictability in stock returns.

It is so widely accepted that stock returns are predictable that it has become a stylized fact. Lettau and Ludvigson (2001) reiterate the broad acceptance that excess returns are predictable by variables such as dividend-price ratios, and earnings-price ratios. This conclusion is hardly restricted to academia. Wilcox (2007) reports that the clear consensus within the investment industry is that prediction based on these ratios is highly useful. As a byproduct of its establishment as a stylized fact, there have been many economic models which can support some degree of return predictability in a general equilibrium setting. Theoretical devices used to do so include consumption smoothing in Balvers, Cosimano, and McDonald (1990), habit formation in Campbell and Cochrane (1999), heterogeneous preferences in Chan and Kogan (2002), and time varying risk preferences in Menzly, Santos, and Veronesi (2004).

Despite this general acceptance, testing predictability has remained a popular research topic. Many have been uneasy with the commonplace application of the standard OLS hypothesis testing. Financial data in general, and return data in particular, are widely known to have extensive econometric complications. There

are several widely recognized characteristics of the covariates used to test return predictability which can cause standard hypothesis tests to over reject a true null. The most well documented and explored of these characteristics are a persistence in the regressor and correlation between innovations in the regressor and returns. More than these technical objections, predictive regressions don't actually predict very well. This has been one of the weightiest arguments against the validity of predictive regressions. Bossaerts and Hillion (1999) find that even the best prediction models have no out-of-sample predictive power. Welch and Goyal (2008) find that the standard predictive variables perform poorly in- and out-of-sample and are outperformed by something as simple as the historical average.

Predictive regression data has other problematic tendencies. The existence of a stochastic time-varying volatility has been widely considered and explored in stock return related literature, though never seriously in this context. As the conditional variance is clearly variable, returns have long been modeled as an ARCH or GARCH process. These models explicitly assume the existence of a constant unconditional variance. In this case, time varying volatility will have no effect on OLS estimation. However, Loretan and Phillips (1994) and Stărică and Granger (2005) find strong evidence against a constant unconditional variance. A time-varying stochastic variance process such as a volatility regime switching model or the model by Heston (1993) are natural alternatives. Cavaliere and Taylor (2007) have shown that, in the presence of stochastic and non-stationary volatility, the standard unit root tests are highly distorted. This immediately implies that standard predictability tests, as well, will be heavily distorted by stochastic volatility in returns.

We provide a new technique for testing predictability that is uniquely suited to all of these characteristics. We hope that this approach will be appealing not only for its effectiveness, but also for its simplicity. We combine a simple instrumental

variable estimator with a simple time change. The instrument directly addresses the inherent endogeneity emerging from the correlation between regressor innovations and returns and persistence. We use the Cauchy estimator, an instrumental variable which requires no additional data. The time change corrects any “poor behavior” on the part of the return distributions. Essentially, since stock returns have a volatility that varies over different time periods, we wait for volatility to reach a certain amount and then add that observation to our sample. So, across all observations there is a constant level of volatility. After this procedure, the regression errors and returns are guaranteed to be independent and normally distributed with a variance of our choosing. This point is worth reemphasizing: regardless of the form of the volatility structure of returns, they will always have a normal distribution. Thus, we have ensured the validity of the standard hypothesis tests.

The remainder of the paper is organized as follows. Section B provides the background and summarizes the main issues relating to return predictability. Section C introduces a novel approach to effectively deal with the various problems affecting the conventional approach. In particular, we introduce a time change to volatility time in order to correct for time-varying stochastic volatilities nonparametrically. In Section D, we subsequently present a new test, which is based on the Cauchy t -ratio for the samples collected after the time change. It is shown that the time-changed Cauchy t -ratio has standard normal limit distribution under truly mild regularity conditions. We show that the use of Cauchy t -ratio allows the predictors to have various statistical anomalies such as near nonstationarity, structural breaks and jumps, among many others. Section E provides Monte Carlo evidence that demonstrates the effects of stochastic volatility on traditional predictability tests. We find that in the presence of these problems, the use of OLS and standard hypothesis testing is wholly inappropriate. These results demonstrate that, using our combination of a time change

and the Cauchy estimator, the data characteristics described above are no longer a factor. Furthermore, if the researcher desires increased power, existing tests which are more powerful and were invalid in the presence of stochastic volatility may be applied to the time-changed data. Next, in Section F we apply the technique directly to actual stock return data. We examine the data set recently used in the paper by Campbell and Yogo (2006). The empirical results are clear: we find no evidence for predictability in stock returns. Section G concludes the paper, and all the proofs are in the Mathematical Appendix.

Throughout the paper, we use $=_d$ and \rightarrow_d to denote equality and convergence in distribution, respectively. To denote a stochastic process Z with time index $s \geq 0$, we use Z_s and $Z(s)$ interchangeably, based on saving space and accentuating the argument.

B. Background and Main Issues

In most general equilibrium models, excess stock returns can not be predictable. The null hypothesis of the unpredictability of stock returns (y_i) has been routinely tested with simple regression

$$y_i = \alpha + \beta x_{i-1} + u_i, \quad (2.1)$$

for $i = 1, \dots, N$, where (x_i) is some covariate which is believed to have some predictive power on the future values of (y_i) . If stock returns are not predictable, clearly we will have that $\alpha = \beta = 0$. The most commonly used covariates are those which make the most economic sense, such as the dividend-price ratio and the earnings-price ratio.

To simplify the subsequent discussions, we let (x_i) be univariate as in most applications, unless specified otherwise. Moreover, we momentarily assume in this section that the constant term α is zero or its nonzero value is already incorporated

into the definition of (y_i) , so that we may concentrate on the slope coefficient β in regression (2.1).¹ It will be explained later which part of the subsequent discussions in this section should be modified when there is the constant term in the regression. In most of the related literature, β is estimated by its ordinary least-squares estimate $\hat{\beta}_N$ and tested using the associated t -ratio which we will denote by $\tau(\hat{\beta}_N)$. Under the standard assumptions, we have

$$\tau(\hat{\beta}_N) \rightarrow_d \mathbb{N}(0, 1)$$

as $N \rightarrow \infty$. It is, however, well documented that certain data characteristics may cause the distribution of the standard t -statistic to be far from standard normal, yielding a substantial bias to a test result relying on standard normal critical values. This will be explained in detail below.

For a sequence (ξ_i) of random vectors, let Z_N be a normalized partial sum process defined for $r \in [0, 1]$ as $Z_N(r) = N^{-1/2} \sum_{i=1}^{[Nr]} \xi_i$, where $[z]$ is the integral part of any real number z . Throughout this section, we say that the invariance principle holds for (ξ_i) if the normalized partial sum process Z_N converges in distribution to a vector Brownian motion on $[0, 1]$ as $N \rightarrow \infty$. The covariance matrix of the limit vector Brownian motion will simply be called the long-run variance of (ξ_i) following the usual convention.

1. Persistence and Endogeneity of Predictors

The covariates, (x_i) , commonly used in the predictive regressions all show strong persistency. This is easily observable and has been well noted by many authors. For instance, see Goyal and Welch (2003) or Torous, Valkanov, and Yan (2004). Indeed,

¹The presence of the constant term α in regression (2.1) may be very important from the statistical point of view, as illustrated by Chen and Deo (2008).

it is routinely modeled as an autoregressive process with the autoregressive parameter that is close to one, or more rigorously as a local-to-unity process. To fix the idea, we let

$$x_i = (1 - c/N)x_{i-1} + v_i \quad (2.2)$$

for some $c \geq 0$. Moreover, we let $\xi_i = (u_i, v_i)'$ and assume that the invariance principle holds for (ξ_i) with the bivariate limit Brownian motion $B = (U, V)'$, whose covariance matrix is given by

$$\Omega = \begin{pmatrix} \omega_u^2 & \omega_{uv} \\ \omega_{uv} & \omega_v^2 \end{pmatrix}.$$

Note that (u_i) is a sequence of martingale differences, and consequently, we may expect $\sum_{i=1}^N u_i^2/N \rightarrow_p \omega_u^2$ to hold under mild conditions.

The asymptotic null distribution of the usual OLS t -ratio $\tau(\hat{\beta}_N)$ can now be easily deduced, which is given by

$$\tau(\hat{\beta}_N) \rightarrow_d \frac{1}{\omega_u} \left(\int_0^1 V_c(r)^2 dr \right)^{-1/2} \int_0^1 V_c(r) dU(r), \quad (2.3)$$

where V_c is an Ornstein-Uhlenbeck process defined as a solution to the stochastic differential equation $dV_c(r) = -cV_c(r)dr + dV(r)$ driven by the limit Brownian motion V . To further analyze the limit distribution in (2.3), we introduce

$$W = U - \frac{\omega_{uv}}{\omega_v^2} V,$$

a Brownian motion independent of V . Note that we have $U = (\omega_{uv}/\omega_v^2)V + W$ by construction.

The limit distribution in (2.3) may then be written as the sum of

$$P = \frac{\omega_{uv}}{\omega_u \omega_v^2} \left(\int_0^1 V_c(r)^2 dr \right)^{-1/2} \int_0^1 V_c(r) dV(r)$$

$$Q = \frac{1}{\omega_u} \left(\int_0^1 V_c(r)^2 dr \right)^{-1/2} \int_0^1 V_c(r) dW(r).$$

It is well known that the distribution of Q is normal, due to the independence of V and W . On the other hand, the distribution of P is essentially that of the t -ratio of the AR coefficient in the near-unit root model, obtained previously by Phillips (1987). The actual null distribution of $\tau(\hat{\beta}_N)$ is a mixture of normal and near-unit root distribution, with the mixing weight given by the long-run correlation coefficient $\rho_{uv} = \omega_{uv}/\omega_u \omega_v$ of (u_i) and (v_i) . We may indeed easily deduce $\tau(\hat{\beta}_N) \rightarrow_d \mathbb{N}(0, 1)$ if $\rho_{uv} = 0$. As $\rho_{uv}^2 \rightarrow 1$, the asymptotic null distribution of $\tau(\hat{\beta}_N)$ diverges from standard normal. See, e.g., Elliot and Stock (1994) for more discussions on the asymptotic null distribution of $\tau(\hat{\beta}_N)$. It is therefore clear that only when $\rho_{uv} = 0$ will the test based on the standard normal distribution be valid. Moreover, we may well expect that the size of the test becomes more distorted as $\rho_{uv}^2 \rightarrow 1$. Perhaps more importantly, it should be noticed that the distribution of the test statistic depends critically on the parameter c . This parameter cannot be estimated consistently since as the sample size N increases, c/N gets smaller at the same rate.

This problem has been clearly demonstrated recently by Campbell and Yogo (2006). In particular, they note that the size distortion is most severe when $c \approx 0$ and $\rho_{uv}^2 \approx 1$. For instance, if there is an exact unit root in the covariate and perfect long-run correlation between the innovations of the covariate and regression errors, they find that the asymptotic size of the one-sided t -test at 5% significance is as large as 46%. The reality of the data does not seem to be far from this worst case scenario. Upon examination, one can easily see that the predictors are highly

persistent. More formally, Campbell and Yogo (2006) report that for the commonly used predictors such as dividend-price ratio and earnings-price ratio, unity lies outside the 95% confidence interval for only ten out of twenty eight data combinations. Even when a unit root may be rejected, the predictors are highly persistent. Of the ten for which unity lies outside the 95% confidence bounds, seven include an autoregressive parameter above 0.95. Moreover, innovations of the predictors seem to be highly correlated with stock returns in the long-run. For instance, the actual sample long-run correlation between differences in the dividend-price ratio and stock returns is -0.98.

2. Nonstationary Stochastic Volatility in Returns

The returns, (y_i) , which would be identical to the regression errors (u_i) under the null of no predictability, are widely believed to have time-varying stochastic volatility. In this subsection, we introduce various nonstationary stochastic volatility models considered in the literature. Following the usual specification for volatility model, we let

$$u_i = \sigma_{i-1}\varepsilon_i, \tag{2.4}$$

where (ε_i) is a martingale difference sequence with respect to filtration (\mathcal{F}_i) such that $\mathbb{E}(\varepsilon_i^2|\mathcal{F}_{i-1}) = 1$ for all $i \geq 1$. Under this specification, we have $\mathbb{E}(u_i^2|\mathcal{F}_{i-1}) = \sigma_{i-1}^2$, and therefore, σ_{i-1}^2 becomes the conditional variance of u_i given information at time $i - 1$, $i \geq 1$. The volatility process (σ_i) is known to be very persistent and at least nearly nonstationary. Indeed, many authors have found that the AR parameter for the volatility process is close to unity under some appropriate functional transformations. See, for instance, Jacquier, Polson, and Rossi (2004), who provide convincing evidence that the log of volatility process follows a near-unit root process for a very wide range

of equity and exchange rate time series. We may then conclude that the true volatility process is highly nonstationary, since it will be the exponential of a near-unit root process.

To accommodate this nonstationarity in volatility, we let

$$\sigma_i = \varpi(z_i) \tag{2.5}$$

with some near-unit root process (z_i) and $\varpi : \mathbb{R} \rightarrow \mathbb{R}_+$. In what follows, we let

$$z_i = (1 - c/N)z_{i-1} + w_i \tag{2.6}$$

for some $c \geq 0$,² and assume that ϖ is asymptotically homogeneous in the sense of Park and Phillips (1999), i.e., $\varpi(\lambda x) \approx \pi(\lambda)\bar{\varpi}(x)$ in x uniformly over any compact subset of \mathbb{R} for all large λ . We call π and $\bar{\varpi}$ respectively the asymptotic order and limit homogeneous function. Loosely, an asymptotic homogeneous function is a function that behaves like a homogeneous function in the limit.

In place of (2.5), we may set

$$\sigma_i = \varpi\left(\frac{i}{N}\right) \tag{2.7}$$

with ϖ being a fixed function or a random function independent of other stochastic components of the model as in Cavaliere (2004) or Cavaliere and Taylor (2007), who studied the unit root test in the presence of stochastic volatility in the innovations. We may also consider the volatility model given by

$$\sigma_i = \varpi\left(\frac{z_i}{\sqrt{N}}\right). \tag{2.8}$$

²We may of course allow the local-to-unity parameter c of (z_i) to be different from that of (v_i) in (2.2). The same $c \geq 0$ is used simply to avoid introducing an additional parameter.

The asymptotic distribution of $\tau(\hat{\beta}_N)$ under the specification of volatility in (2.7) and (2.8) is largely comparable and can be easily obtained from our result based on (2.5).

The theory of regression with errors (u_i) specified as in (2.4) and (2.5) has recently been developed by Chung and Park (2007). Redefine $\xi_i = (\varepsilon_i, v_i, w_i)'$, where (v_i) , (ε_i) and (w_i) are introduced respectively in (2.2), (2.4) and (2.6), and assume that the invariance principle holds for (ξ_i) with the limit Brownian motion denoted by $B = (U, V, W)'$. Also, we define W_c to be the Ornstein-Uhlenbeck process with the mean reversion parameter $c \geq 0$. Then under the additional condition that $\sup_{i \geq 1} \mathbb{E}(|\varepsilon_i|^{2+\epsilon} | \mathcal{F}_{i-1}) < \infty$ a.s. for some $\epsilon > 0$, we have

$$\tau(\hat{\beta}_N) \rightarrow_d \frac{\int_0^1 V_c(r) \bar{\varpi}(W_c(r)) dU(r)}{\left(\int_0^1 \bar{\varpi}(W_c(r))^2 dr \right)^{1/2} \left(\int_0^1 V_c(r)^2 dr \right)^{1/2}}.$$

Under the specification of volatility in (2.7) and (2.8), we have the same result only with the replacement of $\bar{\varpi}(W_c(r))$ by $\varpi(r)$ and by $\varpi(W_c(r))$, respectively.

The limit null distribution of the standard t -ratio $\tau(\hat{\beta}_N)$ is non-normal, even when the innovations of the covariate (v_i) and those of the volatility factor (w_i) are completely independent of the innovations of errors (ε_i) . How far it is away from the standard normal depends on many factors including the volatility function, asymptotic covariances of the innovations and local-to-unity parameters. Given the previous simulation studies by Chung and Park (2007) and Cavaliere and Taylor (2007), we may expect substantial size distortion from using the standard normal critical values in the predictive regression setting.

3. Other Issues

The limiting null distribution of the OLS t -ratio $\tau(\hat{\beta}_N)$ we obtained in (2.3) is not robust with respect to a wide range of other statistical problems in stock return data. In

particular, the distributions are dependent upon the presence of deterministic trends, thick tails in the innovations, jumps and structural breaks, among other things, in the predictive ratios. The existence of any of these problems generally affects the limiting null distributions of $\tau(\hat{\beta}_N)$, and is likely to introduce further size distortion to the test based on standard normal critical values.

The presence of deterministic trends in some predictive ratios, especially in the 1990's, has been widely discussed. The possibility of their having structural breaks in the mean and volatility has also drawn some attentions in the literature. Lettau and Van Nieuwerburgh (2008) focus on structural breaks in the mean of predictive ratios. They show that a small break in the mean can explain some of the common characteristics of predictive regressions. Moreover, we may also infer from Kim, Leybourne, and Newbold (2004), which studies the behavior of unit root tests, that a structural break in the variance of innovation can seriously distort the distribution of the OLS t -ratio in predictability tests. In this paper, we make no claims regarding the existence of time trends and structural breaks, or the lack there of. Instead, we simply point out that if there is a deterministic trend or structural break and a research does use critical values appropriate to the no-trend or no-break case (or vice versa), inference is rendered spurious.

Thick tails in the innovations of predictive ratios would also affect the distribution of the OLS t -ratio. In the context of the unit root test, Ahn, Fotopoulos, and He (2001) show that, for the value of stability index 1.5, the rejection probability of the 5% test is only 0.6% under the null hypothesis. The asymptotic distribution of $\tau(\hat{\beta}_N)$ for this case indeed follows straightforwardly from the invariance principle for stable innovations, for which the reader is referred to Borodin and Ibragimov (1995). Returns also have long been thought to have relatively thick tails. However, we do not use stable distributions to model thick tails in returns. The nonstationary stochastic

volatility in returns also generate thick tails, as shown in Park (2002), and for the purposes of this paper it does not seem to be useful to introduce an additional source of thick tails in returns.

It is worth emphasizing that the size distortion problem of the usual OLS t -ratio we address here is not simply a finite sample phenomenon. Virtually all the previous studies on stock return predictability are based on relatively large samples, with the sample size large enough for the asymptotic theory to provide a good approximation. Further, all the earlier studies further try to bridge the remaining gap between the finite sample distribution and theoretical asymptotic distribution by introducing the local-to-unity formulation in specifying the data generating process for the covariate.

Our previous results for the OLS t -ratio do not change if we include the constant term as in regression (2.1). To get the exact limit null distributions in this case, we simply need to replace V_c by \bar{V}_c , where $\bar{V}_c(r) = V_c(r) - \int_0^1 V_c(s)ds$. In general, the limit null distribution of the OLS t -ratio is further away from the standard normal if the constant term is included. The test based on the standard normal critical values would yield more serious size distortions. See, e.g., Chen and Deo (2008) for some related discussions.

C. A Novel Approach

In this section, we develop a methodology to deal with these problems very effectively. It consists of two separate procedures: a time change and Cauchy estimation. When used together, these two procedures are incredibly robust and well suited to address the problems which we have developed thus far. The time change effectively and non-parametrically corrects for time-varying stochastic volatility in returns. The Cauchy estimator, with a relatively small loss of power, solves all of the potential statistical

problems caused by the covariate nonstationarity with endogeneity, and many other problems that could possibly exist in the covariate such as the presence of deterministic trends, structural breaks, thick tails in the innovations, jumps and outliers. Regardless of the presence of any of these econometric problems, the Cauchy t -ratio of the time-changed data will have an asymptotically standard normal distribution.

3.1 A Change of Chronometer

The approach here relies heavily on the theory of continuous time stochastic processes. Therefore, throughout the paper let the log of stock price, (Y_t) , be a stochastic process in continuous time, which is adapted to a filtration (\mathcal{F}_t) representing the information accumulated up to, and available at, time t . We are interested in testing for the null hypothesis of no predictability of (Y_t) given by

$$\mathbb{E}(dY_t|\mathcal{F}_t) = 0,$$

which implies, in particular, that

$$\mathbb{E}(Y_{t+h} - Y_t|\mathcal{F}_t) = 0$$

for all t and $h > 0$.

Under the null hypothesis, (Y_t) becomes a martingale with respect to the filtration (\mathcal{F}_t) . That is, given all available information up to the current date, t , the best possible prediction in the mean squared sense for any future value of a stock is the current value of that stock. Note that (\mathcal{F}_t) is generally larger than the natural filtration of (Y_t) , i.e., (\mathcal{F}_t) contains more information than is provided only by the realized values of (Y_s) up to time $t > 0$. In our framework, (y_i) in regression (2.1) may be obtained by

$$y_i = Y_{t_i} - Y_{t_{i-1}} \tag{2.9}$$

for some choice of discrete set of time indexes $0 = t_0 < \dots < t_N = T$ over the time interval $[0, T]$. In the subsequent development of our theory, we assume that $x_i = X_{t_i}$ with $t_i \in [0, T]$ for some continuous time stochastic process (X_t) , which is adapted to the filtration (\mathcal{F}_t) .

Now, we let $\langle Y \rangle$ be the quadratic variation of Y , which is defined as

$$\langle Y \rangle_t = \text{plim}_{\pi_t \rightarrow 0} \sum_{k=1}^m (Y_{s_k} - Y_{s_{k-1}})^2,$$

where π_t is the mesh of partition $0 = s_0 < \dots < s_m = t$ of the interval $[0, t]$, i.e., $\pi_t = \max_{1 \leq k \leq m} |s_k - s_{k-1}|$. Our methodology heavily depends upon the following celebrated theorem by Dambis, Dubins and Schwarz.

Lemma 3.2 Let Y be a continuous martingale. Then there exists a standard Brownian motion W such that

$$Y_t = W_{\langle Y \rangle_t},$$

or equivalently,

$$Y_{T_t} = W_t,$$

where T is a time change defined by

$$T_t = \inf_{s \geq 0} \{ \langle Y \rangle_s > t \}.$$

The Brownian motion W is called the DDS Brownian motion of Y .

See, e.g., Revus and Yor (2005) for the proof and more discussions about this result, which is often referred to as the DDS theorem. Note that if $\langle Y \rangle$ is strictly increasing as in most potential applications, T is nothing but the time inverse of $\langle Y \rangle$. See Figure 1 for a demonstration of the change to volatility time. The quadratic variation increases at a regular interval of our choosing, while the calendar time between observations

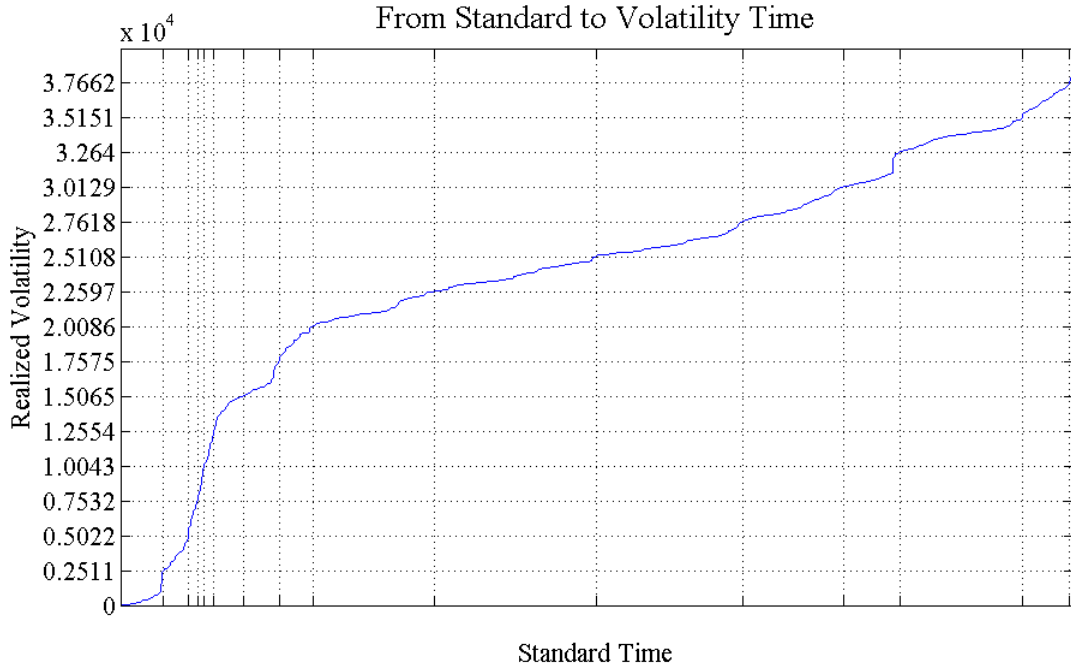


Fig. 1. Calendar versus volatility time. This figure demonstrates the conversion between calendar time and volatility time. Instead of proceeding by one interval of time (i.e., one month), we proceed by one interval of volatility between observations. For illustrative purposes, not all of the random time observations are shown here.

varies greatly. Notice that when the returns are more volatile, which implies that quadratic variation increases quickly, the random time observations are very close to one another. In less volatile periods, when quadratic variation is increasing more slowly, the random time observations are much further apart.

The DDS theorem implies that all continuous martingales are essentially a Brownian motion read according to a different clock. They are merely a time deformation of a Brownian motion. They are different from a Brownian motion only in that a general continuous martingale's quadratic variation is not given by the chronological time clock. On the other hand, this implies that for any continuous martingale, if we use the time clock inversely proportional to its quadratic variation, the martingale

reduces to the standard Brownian motion. We use this important fact to develop a methodology to deal with a general time-varying stochastic volatility in the regression errors (u_i) in regression (2.1).

Under the null hypothesis of no predictability, we have $u_i = y_i$, which is given by (2.9) in our continuous time setup. The regressions errors (u_i) therefore become a sequence of martingale differences, i.e., $\mathbb{E}(u_i|\mathcal{F}_{t_{i-1}}) = 0$ for $i = 1, \dots, N$. However, this provides no information on the conditional and unconditional higher moments. As we explained in the previous section, the presence of heterogeneous volatility makes the use of the standard testing procedure invalid if it is nonstationary.

We may use the DDS theorem to nonparametrically correct for a wide spectrum of nonstationary volatilities in stock returns. To see this, consider a sequence of stopping times

$$0 = T_0 \leq T_\Delta \leq \dots \leq T_{N\Delta} = T$$

for fixed Δ , and let $N = \langle Y \rangle_T / \Delta$. Then we define

$$y_i^* = Y_{T_{i\Delta}} - Y_{T_{(i-1)\Delta}}$$

for $i = 1, \dots, N$. Note that we have

$$Y_{T_{i\Delta}} - Y_{T_{(i-1)\Delta}} = W_{i\Delta} - W_{(i-1)\Delta} =_d \mathbb{N}(0, \Delta),$$

where W is the DDS Brownian motion of Y . Finally, we set $x_{i-1}^* = \sum_s = i-1 X_{T_s\Delta}$ and consider the regression

$$y_i^* = \alpha c_i^* + \beta x_{i-1}^* + u_i^* \tag{2.10}$$

in place of regression (2.1), where $c_i^* = T_{i\Delta} - T_{(i-1)\Delta}$. The variable regressor (c_i^*) is necessary since we have a variable interval between observations rather than the

constant interval of regression (2.1).

In sharp contrast to the original predictive regression (2.1) having errors potentially contaminated with various kinds of nonstationary volatilities, the newly proposed regression (2.10) has errors that are independent normals under the null hypothesis. This extremely useful property is easily attainable. To set up the new regression, instead of obtaining a sample at a fixed time interval, such as monthly, quarterly, or yearly, simply collect the data via a random sampling time based on volatility. The random sampling scheme used here is quite intuitive. We collect the data more often when the market is more volatile, and less often when the market is more stable. The resulting stock returns would then have a constant volatility. Due to the celebrated DDS theorem, if the log of stock price follows a continuous martingale as we assume under the null hypothesis, the stock returns collected in such a manner will always be independent normals.

Of course, the log of stock process Y is not continuously observable. Moreover, its quadratic variation $\langle Y \rangle$ has to be estimated, since it cannot be directly observed. To implement the idea of our time change in practical applications, we should therefore use observations of Y collected at sampling intervals substantially smaller than Δ . For the purposes of this paper, we use the daily stock price data to estimate the quadratic variation of $\langle Y \rangle$ and consider regression (2.10) with Δ , which is the average estimated increase in quadratic variation of $\langle Y \rangle$ over a month, quarter or year. This will be explained in more detail later.

The DDS theorem does not apply if there are jumps in the process. Our previous discussions based on the DDS theorem are therefore no longer true if the log of the stock price process has jumps, which are very likely especially when we use high frequency data. However, our methodology can be easily modified to accommodate for the presence of jumps. One of the easiest remedies is to test for the existence

of jumps in the intervals $[T_{(i-1)\Delta}, T_{i\Delta}]$, $i = 1, \dots, N$. If any of the intervals is suspected of containing jumps, then we may just delete the corresponding time changed observation from the regression. We may of course also try to find the exact timings of jumps and remove them before we apply our methodology. If we can identify the precise location and magnitude of the jump, we can simply remove the jump from the data. For the detection of jumps, the reader is referred to Aït-Sahalia (2004) and Barndorff-Nielsen and Shephard (2006). Clearly, as long as we have finite number of jumps and if the number of jumps is small relatively to N , the effect of discarding the stock returns over the intervals with jumps would not be detrimental. The same remedy may also be applicable for other types of discontinuities, such as structural breaks, in the sample path.

1. Efficient Tests

Once errors in regression (2.1) have been made into independent standard normally distributed observations, the researcher may apply some appropriate efficient tests. The tests we examine here are the Bonferroni based Q -test of Campbell and Yogo (2006) and the Restricted Maximum Likelihood (REML) based test of Chen and Deo (2008). These two tests have been shown to be highly efficient in the case in which returns and regressor innovations are normally distributed and independent. This convenient form is guaranteed by the use of volatility time.

Campbell and Yogo's (2006) test is efficient and asymptotically has the correct size. Since c cannot be consistently estimated, the test relies on placing confidence intervals on c . Then, using the endpoints of this set, confidence bounds may be placed on β . Suppose that a researcher were interested in a 5% test. The researcher could construct a 2.5% confidence interval on c . For each value in this interval, the researcher could then construct a 2.5% interval on β given the value of c . From

Bonferroni, we know that such a confidence interval will have a coverage of at least 5%. The basic asymptotics developed by Campbell and Yogo (2006) are based on a simple standard normal error setup. Thus, the time-changed data should be ideal for optimal use of this estimator.

Chen and Deo (2008) show that a major source of distortion of test statistics in predictive regression is actually from the intercept parameter. Their REML based likelihood ratio test achieves roughly half as much bias as OLS estimates for nearly unit root processes while suffering no loss in efficiency. For now, let η be the autoregressive parameter and let X_c and Y_c be the mean corrected independent and dependent variables, respectively. From Chen and Deo (2008), we have that the REML estimates are given by

$$\hat{\eta}_{\text{REML}} = \arg \min_{\eta} \left\{ n \log \left(\frac{1 + \eta}{(n - 1)(1 - \eta) + 2} \right) \right\},$$

$$(\hat{\rho}_{uv, \text{REML}}, \hat{\beta}_{\text{REML}}) = \begin{pmatrix} 1 & 0 \\ \hat{\eta} & 1 \end{pmatrix} (X_c' X_c)^{-1} X_c' Y_c.$$

Similarly to the OLS estimate, the REML estimate of β is critically dependent on the bias in η . Of course, the structure of the likelihood function that forms the basis of this estimation technique is critically dependent on the distribution. This makes a pre-test time change especially convenient for this estimator so that we know precisely the distribution of the structural error.

From this subsection, the benefits of the time change may easily be seen. Likewise, the results in Tables I and II show results which are much more inline with what we economically would expect. The data used here is described more fully in the empirics in Section 6. At no frequency and for no covariate is the so called predictive ratio a significant predictor of future returns. For the Bonferroni type test

Table I. Estimates based on Campbell and Yogo's (2006) efficient test of predictability performed on the time-changed data.

Frequencies	Covariate	$\hat{\beta}_{CY}$	t -test	Q -test
Monthly	DP	0.003	[-0.016, 0.011]	[-0.012, 0.017]
	EP	0.012	[-0.005, 0.018]	[-0.002, 0.023]
Quarterly	DP	0.008	[-0.049, 0.031]	[-0.039, 0.045]
	EP	0.039	[-0.013, 0.056]	[-0.006, 0.067]
Annual	DP	0.033	[-0.167, 0.123]	[-0.136, 0.161]
	EP	0.185	[-0.030, 0.261]	[-0.013, 0.309]

Table II. Estimates for Chen and Deo's (2008) Restricted Maximum Likelihood (REML) test performed on the time-changed data.

Frequencies	Covariate	$\hat{\beta}_{REML}$	REML Statistic	p -value
Monthly	DP	0.2772	0.5343	0.4648
	EP	0.5814	2.1317	0.1443
Quarterly	DP	0.9350	0.6282	0.4280
	EP	1.8365	2.1977	0.1382
Annual	DP	3.0062	0.3966	0.5289
	EP	6.4617	1.8428	0.1746

of Campbell and Yogo (2006), the right most columns display the 90% confidence interval for each estimate for the t - and Q -tests. In all cases, the null hypothesis of $\beta = 0$ lies within the confidence intervals. For bias reduced likelihood based test of Chen and Deo (2008), the estimates are not significant at any of the conventional levels.

2. A Robust Test - The Cauchy Estimator

Now, using regression (2.10), the errors are guaranteed to be well behaved. As previously stated, it may then be acceptable to apply predictability tests based on Campbell and Yogo (2006) or Chen and Deo (2008). However, we provide an effective way to deal with a wide variety of problems in the covariate of our predictive regression which may remain despite the time change, including persistence and endogeneity and the other data anomalies we discussed in the previous section. In fact, this section demonstrates how to properly address the issues that have been discussed in virtually every previous paper on predictive regressions. To convey the main idea, we consider regression (2.1) with no constant term, i.e., $\alpha = 0$, and introduce the Cauchy estimator $\tilde{\beta}_N$ for β , which is given by

$$\tilde{\beta}_N = \left(\sum_{i=1}^N |x_{i-1}| \right)^{-1} \sum_{i=1}^N \text{sgn}(x_{i-1}) y_i,$$

where $\text{sgn}(\cdot)$ is the sign function defined as $\text{sgn}(x) = 1$ if $x \geq 0$ and $\text{sgn}(x) = -1$ if $x < 0$. Clearly, $\tilde{\beta}_N$ is nothing but the IV estimator using $\text{sgn}(x_{i-1})$ as an instrument. This estimator was first proposed by Cauchy in 1836 and hence is referred to as the “Cauchy estimator”. It has more recently brought some attention in the econometric literature. See So and Shin (1999) and Chang (2002) for the use of Cauchy estimator to test for a unit root. We will show that this estimator will be robust to a truly wide

variety of data characteristics.

The associated Cauchy t -ratio $\tau(\tilde{\beta}_N)$ for β is given by

$$\tau(\tilde{\beta}_N) = \frac{\tilde{\beta}_N}{s(\tilde{\beta}_N)},$$

where $s(\tilde{\beta}_N)$ is the standard error of the Cauchy estimator $\tilde{\beta}_N$, which is given by

$$s(\tilde{\beta}_N) = \hat{\sigma}_N \sqrt{N} \left(\sum_{i=1}^N |x_{i-1}| \right)^{-1}$$

with any consistent estimator $\hat{\sigma}_N^2$ for the asymptotic variance σ^2 of the regression errors (u_i) .³

To develop the asymptotic theory for the Cauchy estimator $\tilde{\beta}_N$ and its t -ratio $\tau(\tilde{\beta}_N)$, we introduce

Assumption 3.1 Let (u_i, \mathcal{F}_i) be a martingale difference sequence such that

- (a) $\frac{1}{N} \sum_{i=1}^N \mathbb{E}(u_i^2 | \mathcal{F}_{i-1}) \rightarrow_p \sigma^2$, and
- (b) $\frac{1}{N} \sum_{i=1}^N \mathbb{E}\left(u_i^2 1\{|u_i| \geq \epsilon \sqrt{N}\} \middle| \mathcal{F}_{i-1}\right) \rightarrow_p 0$ for any $\epsilon > 0$,

as $N \rightarrow \infty$.

The conditions in Assumption 3.1 are not stringent, and are required for the central limit theory to be applicable for (u_i) . See, e.g., Hall and Heyde (1980). The condition in (a) is satisfied for a wide class of martingale sequences. It actually allows for stationary stochastic volatility to be present in (u_i) . If we write $u_i = \sigma_{i-1} \varepsilon_i$ as before, where (σ_i^2) is ergodic and stationary with $\mathbb{E}\sigma_i^2 = \sigma^2$, then the condition obviously holds. The condition in (b) is the conditional version of the usual Lindeberg condition.

³Under very mild conditions, the OLS residuals (\hat{u}_i) , $u_i = y_i - x_i \hat{\beta}_N$, as well as the Cauchy residuals (\tilde{u}_i) , $u_i = y_i - x_i \tilde{\beta}_N$ are consistent. This will be made clear in our subsequent development of our theory.

It is met if we set $\sup_{i \geq 1} \mathbb{E}(|u_i|^{2+\epsilon} | \mathcal{F}_{i-1}) < \infty$ a.s. with some $\epsilon > 0$, as is often assumed in the literature. Consequently, Assumption 3.1 is expected to hold for a general class of martingale difference sequences. It should, however, be emphasized that it does not hold if the nonstationary volatility introduced in Section 2.2 is present, and our subsequent results in this section do not apply. This case will be fully addressed in the next subsection.

Assumption 3.2 There exists a sequence κ_N of numbers such that

$$\left(\kappa_N^{-1} \sum_{i=1}^N |x_i| \right)^{-1} = O_p(1)$$

for all large N .

The required condition in Assumption 3.2 is extremely mild, and should hold for a truly wide variety of the predictor (x_i) . If (x_i) is a nonconstant, stationary and ergodic time series, then the condition is satisfied with $\kappa_N = N$. If, on the other hand, (x_i) has a near-to-unit root and if its innovations satisfy the invariance principle, then we have

$$N^{-3/2} \sum_{i=1}^N |x_{i-1}| \rightarrow_d \int_0^1 |V_c(r)| dr,$$

where V_c is the Ornstein-Uhlenbeck process introduced in the previous section. Therefore, the condition holds for $\kappa_N = N^{3/2}$. We may also allow for fat-tailed innovations of (x_i) . In this case, the required condition is satisfied under general regularity conditions with $\kappa_N = N^{1+1/\alpha} \ell(N)$ for $0 < \alpha < 2$, where α is the stability index and ℓ is a function that is slowly varying at infinity. See, e.g., Borodin and Ibragimov (1995) for more details on the invariance principle for stable innovations. More importantly, the condition in Assumption 3.2 allows for various kinds of data anomalies in (x_i) . The jumps and structural breaks in (x_i) are generally permitted, as long as their numbers

are finite.

Now we may readily deduce the asymptotics of the Cauchy estimator $\tilde{\beta}_N$ and the Cauchy t -ratio $\tau(\tilde{\beta}_N)$.

Lemma 3.1

- (a) If Assumptions 3.1 and 3.2 hold, then $\tilde{\beta}_N = \beta + O_p(N^{1/2}/\kappa_N)$ for all large N .
- (b) If Assumption 3.1 holds and $\beta = 0$, then $\tau(\tilde{\beta}_N) \rightarrow_d \mathbb{N}(0, 1)$ as $N \rightarrow \infty$.

The Cauchy estimator $\tilde{\beta}_N$ is therefore generally consistent. For each of the cases where (x_i) is stationary, nearly-nonstationary and nonstationary with fat-tailed innovations, its convergence rate is given by $N^{1/2}$, N and $N^{(2+\alpha)/2\alpha}\ell(N)$. This gives us two highly useful properties of the Cauchy estimator $\tilde{\beta}_N$. First, its convergence rate is generally the same as the OLS estimator $\hat{\beta}_N$. Second, the Cauchy t -ratio has the standard normal limit distribution. Note that we only impose the conditions necessary for the central limit theory to hold for (u_i) . It is important to notice in particular that we do not require any regularity conditions on (x_i) . This is truly remarkable. In light of this fact, the Cauchy t -ratio is well suited to test for return predictability.

In the classical regression setting, the Cauchy estimator is less efficient than the OLS estimator. If (x_i) is stationary and ergodic, then the asymptotic variance of the former is $\sigma^2(\mathbb{E}|x_i|)^{-2}$, while that of the latter is $\sigma^2(\mathbb{E}x_i^2)^{-1}$, if we assume that they exist. We may therefore easily see that the OLS estimator has a smaller variance than the Cauchy estimator due to the Jensen's inequality. If, on the other hand, (x_i) is a near-unit root process having innovations (v_i) that are asymptotically independent of (u_i) , then the limit distributions of the Cauchy and the OLS estimators are normal mixtures with mixing variates given respectively by $\sigma(\int_0^1 |V_c(r)|dr)^{-1}$ and $\sigma(\int_0^1 V_c(r)^2 dr)^{-1/2}$. It can therefore easily be deduced from the Cauchy-Schwarz inequality

ity $\int_0^1 |V_c(r)|dr \leq (\int_0^1 V_c(r)^2 dr)^{1/2}$ a.s. that the OLS estimator is more efficient than the Cauchy estimator. So, in these highly counterfactual circumstances, the OLS estimator is preferable.

However, in the context of predictive regressions, none of the above standard comparisons between the OLS and Cauchy estimators is applicable. In particular, the relative efficiency of the OLS estimator does not apply to the case where we have persistence and endogeneity considered in Section 2.1. The asymptotic distribution of the OLS estimator is generally biased and skewed, as well as non-normal, whereas that of the Cauchy estimator is normal even in this case. Therefore, the strict comparison based on their asymptotic variances cannot be made. Their relative efficiency will be dependent on the asymptotic correlation between the regression errors (u_i) and the innovations of predictive ratios (v_i), and also on the realization of the predictive ratios (x_i) themselves. Moreover, as discussed in Section 2.3, the OLS estimator is non-robust to other aberrant data characteristics such as deterministic trends, structural breaks and thick tails. In contrast, the Cauchy estimator is robust against these and many other potential problems in the data.

We have more compelling reasons why the Cauchy t -ratio $\tau(\tilde{\beta}_N)$ is preferred to the OLS t -ratio $\tau(\hat{\beta}_N)$ in testing for return predictability. First, the asymptotic null distribution of $\tau(\hat{\beta}_N)$ is generally biased, skewed and has a tail thicker than that of normal distribution. This contrasts to $\tau(\tilde{\beta}_N)$, whose asymptotic null distribution is standard normal. Therefore, against the general alternative of predictability, the Cauchy t -ratio is expected to be more powerful than the OLS t -ratio. Second, more importantly, the asymptotic null distribution of the OLS t -ratio is dependent upon the local-to-unity parameter c of the predictive ratio, which is assumed to be a near-unit root process. Strictly speaking, this nuisance parameter dependency makes the OLS t -ratio unusable. Campbell and Yogo (2006) circumvent this problem by constructing

a confidence interval for c and using a Bonferroni type inequality.

The robustness of the Cauchy estimator, as with all instrumental variable estimation, comes with a loss of some power. In Figure 2, we compare the powers of the Bonferroni-type Q -test of Campbell and Yogo (2006) and the REML test of Chen and Deo (2008) with the Cauchy t -test for some selected values of c and ρ_{uv} when regressors and structural errors are well behaved, in that they follow the requirements of Campbell and Yogo (2006). The Q -test generally out performs the Cauchy t -test in terms of power. The REML test outperforms the Cauchy t - and the Bonferroni Q -tests in most instances. The difference is most striking. Though, the gap in the powers of the Cauchy t -test and the other tests diminishes as $\rho_{uv} \rightarrow 0$. This is well expected, since the asymptotic null distribution of the Bonferroni-type tests approaches the standard normal distribution as $\rho_{uv} \rightarrow 0$. In fact, the power curves of all the tests considered here become roughly comparable when $\rho_{uv} = 0$.

It is important to note that in this subsection, we maintain that the regression errors (u_i) are martingale differences whose volatilities are asymptotically constant. Under this circumstance, we clearly demonstrate that the simple Cauchy t -test alone may effectively deal with the persistence and endogeneity of predictive ratios, the issue specifically considered in the vast majority of previous literature in the predictability of stock returns. However, the Cauchy t -test, as well as all the other existing tests, is not applicable if the stock returns have nonstationary volatilities considered in Section 2.2. In this case, the results in Lemma 3.1 are not valid, and in particular, the limit distribution of the Cauchy t -ratio $\tau(\tilde{\beta}_N)$ is not normal. Therefore, we may no longer use the normal critical values. Perhaps the most valuable contribution of this paper is that, given that the data has already been observed in volatility time, we may maintain these ordinarily strenuous restrictions without loss of generality.

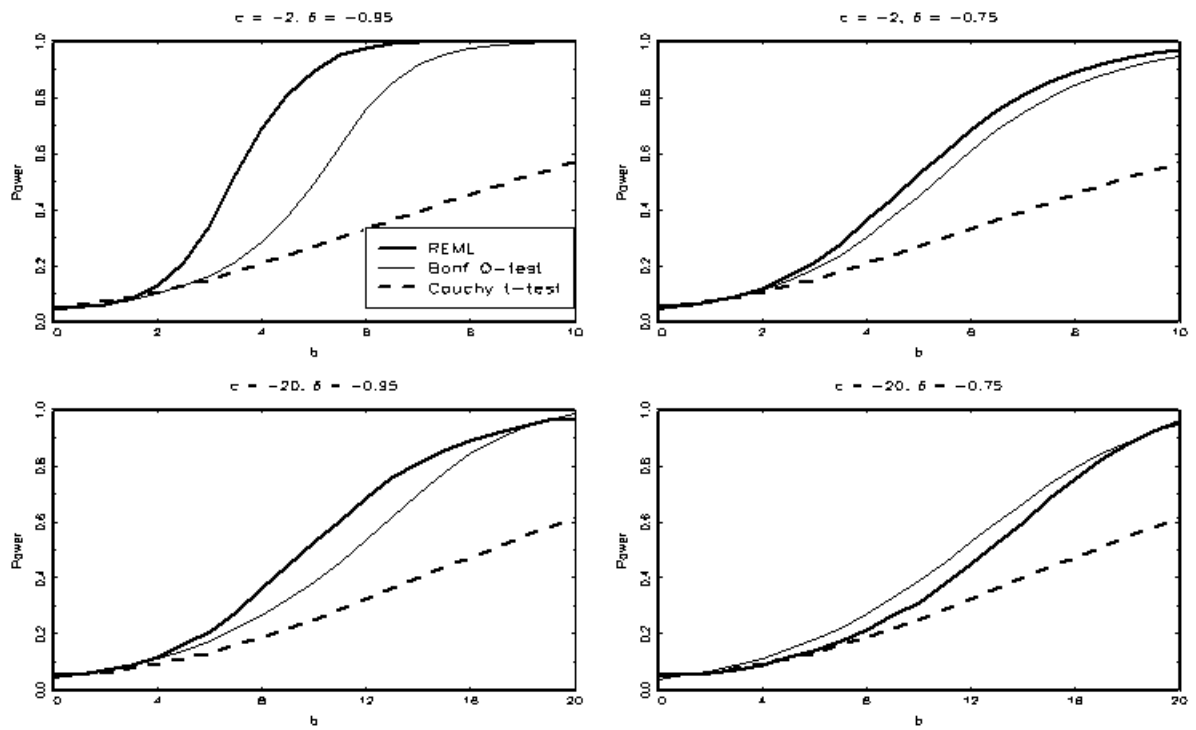


Fig. 2. Power comparison. A comparison of local asymptotic power for the Cauchy t -test, the infeasible t -test, and the Bonferroni Q -test. Here, let $b = T\beta$ to account for the sample size. This figure is generated using 20,000 replications of a 500 observation sample.

D. A New Test and Its Statistical Theory

Our test is based on regression (2.10) consisting of the samples collected from the process Y using the chronometer running at the speed inversely proportional to the quadratic variation $\langle Y \rangle$ of Y . To implement our methodology for practical applications, we therefore need to observe Y and its quadratic variation $\langle Y \rangle$. However, the observations of Y are made only discretely in time. As a result, $\langle Y \rangle$ is not directly observable. We need to estimate the quadratic variation using a discrete set of observations. In this section, we assume that the samples $(Y_{i\delta})$, $i = 1, \dots, n$ are observed over the time interval $[0, T]$ with $T = n\delta$. Note that the size n of the available sample is in general different from the number N of observations used to run regression (2.10). In our applications reported in the paper, we use n to denote the number of daily observations, while N refers to the sample size number of monthly, quarterly or yearly observations.

1. A New Test

To employ our method, we first need to estimate the quadratic variation $\langle Y \rangle$ of Y . The natural estimate for $\langle Y \rangle$ is

$$\langle Y \rangle_t^\delta = \sum_{i=1}^m (Y_{i\delta} - Y_{(i-1)\delta})^2$$

for $(m-1)\delta \leq t < m\delta$, $m = 1, \dots, n$, which is often referred to as the realized variance. Subsequently, we let

$$T_t^\delta = \inf_{s \geq 0} \{ \langle Y \rangle_s^\delta > t \},$$

i.e., the required time change defined from the realized variance process $\langle Y \rangle^\delta$. For a fixed $\Delta > 0$, we let

$$\begin{aligned} y_i^{*\delta} &= Y_{T_{i\Delta}^\delta} - Y_{T_{(i-1)\Delta}^\delta}, \\ c_i^{*\delta} &= T_{i\Delta}^\delta - T_{(i-1)\Delta}^\delta, \end{aligned}$$

and $x_{i-1}^{*\delta} = X_{T_{(i-1)\Delta}^\delta}$. Then we consider the regression

$$y_i^{*\delta} = \alpha c_i^{*\delta} + \beta x_{i-1}^{*\delta} + u_i^{*\delta}, \quad (2.11)$$

which corresponds to regression (2.10) in the previous section.

For the actual test of no predictability in stock returns, we use the excess returns to adjust for the nonzero constant term in regression. This is in accordance with virtually all the previous studies reported in the literature. All those studies found that the constant term, after the adjustment, is insignificant. Therefore, assuming $\alpha = 0$ in regression (2.11), we first consider the test of hypothesis $\beta = 0$. The test of whether the mean of the excessive returns is zero will be introduced later in this section. As we will explain in more detail later, our test results are consistent with earlier findings by others, and strongly support that the excess returns have a mean of zero.

To test for the hypothesis $\beta = 0$, we consider the regression

$$y_i^{*\delta} = \beta x_{i-1}^{*\delta} + u_i^{*\delta} \quad (2.12)$$

and use the Cauchy t -ratio $\tau(\tilde{\beta}_N^{*\delta})$ for β , which is given by

$$\tau(\tilde{\beta}_N^{*\delta}) = \frac{\tilde{\beta}_N^{*\delta}}{s(\tilde{\beta}_N^{*\delta})}.$$

As before, $\tilde{\beta}_N^{*\delta}$ is the Cauchy estimator for β , i.e.,

$$\tilde{\beta}_N^{*\delta} = \left(\sum_{i=1}^N |x_{i-1}^{*\delta}| \right)^{-1} \sum_{i=1}^N \text{sgn}(x_{i-1}^{*\delta}) y_i^{*\delta},$$

and $s(\tilde{\beta}_N^{*\delta})$ is the standard error of the Cauchy estimator $\tilde{\beta}_N^{*\delta}$ that is given as

$$s(\tilde{\beta}_N^{*\delta}) = \hat{\sigma}_N^* \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N |x_{i-1}^{*\delta}| \right)^{-1}$$

with any consistent estimator $\hat{\sigma}_N^{*2}$ for the variance σ^{*2} of (u_i^*) . Note that, by construction, we know that $\sigma^{*2} = \Delta$ if $u_i^* = y_i^*$ for all $i = 1, \dots, N$. We will show that the Cauchy t -ratio $\tau(\tilde{\beta}_N^{*\delta})$ has the standard normal null limiting distribution under very general regularity conditions. The usual normal critical values can therefore be used for the test.

In implementing our test, we pay a particular attention to maximizing the finite sample power. As we mentioned earlier, much of the existing literature finds at least some and usually strong evidence of return predictability in stocks. This is in sharp contrast with our results, which unambiguously finds no support for predictability at all horizons. To strengthen our conclusion, it is therefore particularly important to increase the finite sample power of our test. First, if the mean of excess returns is zero, then the nonzero mean (in the stationary case) or large starting value (in the nonstationary case) of the predictor may decrease the finite sample power of our test. To get rid of the dependency of our test on the nonzero mean or the initial value of the predictor, we may use

$$x_{i-1}^{*\delta} - x_0 \quad \text{or} \quad x_{i-1}^{*\delta} - \frac{1}{i-1} \sum_{j=1}^{i-1} x_{j-1}^{*\delta}$$

or any other transforms relying only on the past values of the predictor, instead of $(x_{i-1}^{*\delta})$ in regression (2.12). Second, the consistent estimator $\hat{\sigma}_N^{*2}$ for the variance of

(u_i^*) is obtained by the OLS regression (2.12) with $(y_i^{*\delta})$ replaced by

$$y_i^{*\delta} - \tilde{\alpha}_N^{*\delta} c_i^{*\delta}$$

with $\tilde{\alpha}_N^{*\delta}$ introduced subsequently below, in addition to the transformation of the predictor mentioned above. We use the demeaned regressand since the nonzero mean in the excess returns, though it is not significantly different from zero, may inflate the estimate of the error variance and falsely lead to nonrejection of no predictability. The OLS method, rather than the IV method, is used to minimize the variance estimate. This also makes it easier to reject the null of no predictability.

To see whether the excess returns have mean zero, we test $\alpha = 0$ in the regression

$$y_i^{*\delta} = \alpha c_i^{*\delta} + u_i^{*\delta}, \quad (2.13)$$

where $c_i^{*\delta} = T_{i\Delta}^\delta - T_{(i-1)\Delta}^\delta$. For regression (2.13), we use the IV approach using the lagged regressor as an instrument, which yields the estimator

$$\tilde{\alpha}_N^{*\delta} = \left(\sum_{i=1}^N c_i^{*\delta} c_{i-1}^{*\delta} \right)^{-1} \sum_{i=1}^N c_{i-1}^{*\delta} y_i^{*\delta}.$$

Our test for the hypothesis $\alpha = 0$ is based on the associated t -ratio, denoted similarly as before by $\tau(\tilde{\alpha}_N^{*\delta})$. To be consistent with our test for β , we use the OLS method to obtain the error variance estimate. It is quite clear that the asymptotic distribution of the IV t -ratio $\tau(\tilde{\alpha}_N^{*\delta})$ also has the standard normal distribution under general regularity conditions. However, we will not formally develop the asymptotics for $\tau(\tilde{\alpha}_N^{*\delta})$ here. They are relatively straightforward, given the asymptotics for $\tau(\tilde{\beta}_N^{*\delta})$ that are presented in this paper.

2. Asymptotic Theory

Our asymptotics require $\delta \rightarrow 0$ and $n \rightarrow \infty$, as well as $N \rightarrow \infty$. Not only must the number of samples increase, but also the maximal frequency of available observations must increase as well. We develop the null asymptotics in this section, and therefore, the null hypothesis of no predictability will be maintained throughout the section. Furthermore, we assume that the jumps and other discontinuities in the path of Y have already been addressed, as explained earlier in Section 3.2. Under the convention, Y becomes a continuous martingale. Intuitively, it is clear that realized variance $\langle Y \rangle^\delta$ gets close to the actual quadratic variation $\langle Y \rangle$, and likewise, a time change based on realized variance is close to the ideal time change, as $\delta \rightarrow 0$. Therefore, if δ is sufficiently small, then we can use random time observations based on realized variance as a valid proxy for random time observations based on quadratic variation. They will be close enough, in fact, that the same asymptotic theory will hold, under very general regularity conditions. To proceed formally, we first introduce some technical conditions. For brevity, we denote $\langle Y \rangle_s^t = \langle Y \rangle_t - \langle Y \rangle_s$ in what follows.

Assumption 4.1 For any $0 < s < t$, there exists some $\kappa > 0$ such that

$$\mathbb{E} \left[(Y_t - Y_s)^2 - \langle Y \rangle_s^t \right]^2 \leq c |t - s|^{1+\kappa}$$

where $c > 0$ is some constant independent of $s > 0$ and, $t > 0$.

A large class of continuous martingales satisfy this assumption. The reader is referred to Park (2008) for more discussions on this assumption. Roughly, this assumption bounds the variance of the error when estimating quadratic variation with realized variance.

Assumption 4.2 Let $\delta \rightarrow 0$ and $n \rightarrow \infty$ such that $n\delta^{1+\kappa} \rightarrow 0$ for the κ introduced above, and such that

$$N \leq \frac{c\Delta}{(n\delta^{1+\kappa})^{1/2} [\log(1/n\delta^{1+\kappa}) + \log(N\Delta)]^2},$$

for some constant $c > 0$.

Essentially, the conditions in Assumption 4.2 require that the frequency of the data increases quickly and that the number of observations increases relatively slowly. They also provide an upper bound for the number of time-changed observations.

It is shown in Park (2008) that

Lemma 4.1 Under Assumptions 4.1 and 4.2, we have

$$\sup_{t \leq n\delta} |\langle Y \rangle_t^\delta - \langle Y \rangle_t| = O_p((n\delta^{1+\kappa})^{1/2}),$$

and

$$\max_{1 \leq i \leq N} |y_i^{*\delta} - y_i^*| = o_p(N^{-1/2})$$

for all large N .

The conditions in Assumptions 4.1 and 4.2 are therefore sufficient to ensure that realized variance uniformly and consistently estimates quadratic variation. As a result, the samples collected using a time change based on realized variance get close to those by the ideal time change, and the errors incurred by using realized variance rather than quadratic variation become negligible asymptotically. Below we show that the t -ratios derived from the estimated time change are asymptotically distributionally equivalent to the case in which the true time change is known.

We may readily deduce from Lemma 4.2 that

Theorem 4.2 Under Assumptions 4.1 and 4.2, we have

$$\tau(\tilde{\beta}_N^{*\delta}) \rightarrow_d \mathbb{N}(0, 1)$$

as $N \rightarrow \infty$.

In Theorem 4.2, we formally establish that the limit distribution of the Cauchy t -ratio is standard normal. Note that we require only very minimal conditions here. We allow the underlying process Y to be a very general martingale having a variety of possibly nonstationary stochastic volatilities. Furthermore, we impose no conditions on the predictor. The asymptotic normality of the Cauchy t -ratio holds regardless of any statistical anomalies in covariates including nonstationarity, fat-tailed innovations, structural breaks and jumps.

Now consider the case where $\beta \neq 0$. In this case, we assume

Assumption 4.3 Let

- (a) $\frac{1}{N} \sum_{i=1}^N \mathbb{E}(u_i^{*\delta 2}) < \infty$, and
- (b) $\sum_{i=1}^N |x_i^{*\delta}|$ is of order κ_N with $\kappa_N/\sqrt{N} \rightarrow \infty$.

Then we have

Proposition 4.3 Let $\beta \neq 0$. Under Assumption 4.3, $\tau(\tilde{\beta}_N^{*\delta})$ diverges at the rate of κ_N/\sqrt{N} as $N \rightarrow \infty$.

The conditions in Assumption 4.3 are very mild. Therefore, Proposition 4.3 shows that the test based on the Cauchy t -ratio $\tau(\tilde{\beta}_N^{*\delta})$ is generally consistent, and have unit power against the alternative $\beta \neq 0$.

Though condition (a) is weak and satisfied widely, it is not followed from any of our previous assumptions. Note that $(u_i^{*\delta})$ may no longer be close to a sequence of

independent standard normals, since the time change is obtained from the quadratic variation of Y and now $\beta \neq 0$. If the process X generating (x_i) is of bounded variation, the quadratic variation of Y is the same as that of process generating the regression error (u_i) . Therefore, if we assume that the generating process of (u_i) is a continuous martingale, then $(u_i^{*\delta})$ reduces to an approximate iid sequence of normals. However, the generating process X of covariate (x_i) may not be of bounded variation. In this case, the quadratic variation of Y becomes in general different from that of the generating process of (u_i) .

It seems obvious that condition (b) holds for all applications related to return predictability. The condition is easy to check, since $(x_i^{*\delta})$ is observed. As we mentioned earlier, for each of the cases where $(x_i^{*\delta})$ is stationary, nearly-nonstationary and nonstationary with fat-tailed innovations, its convergence rate is given by N , $N^{3/2}$ and $N^{(3+\alpha)/2\alpha}\ell(N)$. Therefore, it is easy to see that the condition is met in all these cases.

E. Monte Carlo Simulations

In this section, we begin to examine the effects that nonstationary volatility can have on existing tests as well as on our proposed test. We start with a general model encompassing previous econometric models of predictive regressions. We then examine six different nonstationary volatility structures. First, we introduce a simple structural break; that is, a single change in the volatility of errors and innovations. We also consider a regime switching model with high and low volatility regimes. Finally, we consider the more complex stochastic volatility models of exponential stochastic volatility and the Heston model.

Since our methodology and its theory are developed in a continuous time framework, our Monte Carlo simulations are conducted based on continuous time models.

Therefore, we specify the return and predictor processes, Y and X . Further, we assume that these processes are observed at δ -intervals. Throughout our simulations, we set $\delta = 1/250$ so that $n = T/\delta$, implying that the daily observations are available for both Y and X . For the process Y , we consider

$$dY_t = \sigma_t dW_t, \quad (2.14)$$

where W is a standard Brownian motion. In the typical model of predictability, σ_t is a constant. For more complex and realistic models the exact structure of volatility will vary from model to model. Later in this section, we discuss the volatility processes σ_t that we analyze.

For X , we use

$$dX_t = \kappa(\mu - X_t)dt + \sigma_t dV_t \quad (2.15)$$

with V a Brownian motion, i.e., an Ornstein-Uhlenbeck process. The Brownian motions W and V have correlation ρ_{uv} . For the purposes of simulation, we let $\rho = -0.98$, or the approximate empirical correlation of returns with differenced predictive ratios. Here, we always assume that $\mu = 0$.

For a direct comparison between the more general format we use here, and the existing literature, we introduce a second distinct interval. While δ is a frequency at which the observations are available (here we use daily), let D be the frequency of observations at which the regression is run (generally monthly, quarterly, or yearly) and $N = T/D$. If X is observed at discrete D -intervals, then the corresponding AR regression of this process becomes

$$X_{iD} = (1 + e^{\kappa D})X_{(i-1)D} + v_{iD} \quad (2.16)$$

for $i = 1, \dots, N$ where $v_t \sim \mathbb{N}(0, \sigma^2(1 - e^{2\kappa D} - 1)/2\kappa)$. Thus, this model nests the

discrete time formulations of previous studies like Campbell and Yogo (2006) and Chen and Deo (2008), which instead use a constant value of σ_t . We set κ such that $e^{-\kappa D} \approx 1 - \kappa D = 1 - c/N$. Indeed, if κ is selected in this way and $\sigma_t = 1$ for all t , then (2.16) gives the exact model used by Campbell and Yogo (2006) and Chen and Deo (2008).

Turning now to the various volatility structures, we first consider the most simple case of nonstationary volatility: a single break in the standard deviation of errors and innovations. The return and predictor processes are defined as above, but the volatility process follows

$$\sigma_t = \begin{cases} \sigma_1 & \text{for } t \in [0, T/2] \\ \sigma_2 & \text{for } t \in (T/2, T]. \end{cases}$$

The existence of structural breaks in return volatility has been widely hypothesized. Furthermore, as has been shown previously in Kim, Leybourne, and Newbold (2004) as well as Cavaliere and Taylor (2007), unit root tests can be highly affected by persistent changes in volatility (i.e., nonstationary volatility), which directly implies that standard tests of predictability will also be affected. For this experiment, we consider a three sizes of breaks ($\sigma_1 = 1.5, 3$, or 4.5 and $\sigma_2 = 1$) in order to isolate the effects specific to the change in volatility. These standard deviation ratios are typical of the Kim, Leybourne, and Newbold (2004).

Next, we turn from deterministic changes in volatility to stochastic changes. We address three separate stochastic volatility models. The first stochastic volatility model examined is a simple regime shift in volatility with two possible regimes. Again, this process follows equations (2.14) and (2.15) with σ_t defined in the following way. Let S_t be the indicator function of the current state of the world. In the low volatility state, $S_t = 0$, while in the high volatility state, $S_t = 1$. As in Veronesi (1999), the

process will or will not transition to another volatility state in any given time period with probabilities given by

$$\mathbb{P}[S_{t+h} = 0 | S_t = 0] = p_L h,$$

$$\mathbb{P}[S_{t+h} = 1 | S_t = 1] = p_H h,$$

and so

$$\sigma_t = (1 - S_t)\sigma_L + S_t\sigma_H.$$

Schaller and van Norden (1997) find that for monthly stock return data the probability of remaining in a low volatility regime to be 0.9908 and a 0.9411 probability of staying in a high volatility regime, with low and high regime standard deviations $\sigma_L = 0.0392$, and $\sigma_H = 0.1180$, respectively. Correspondingly, we assign these values to standard deviation and the transition probabilities to $p_L = 0.9996$ and $p_H = 0.9976$ for our daily data.

Lastly, we consider two additional, more sophisticated, models of stochastic volatility. Specifically, we consider exponential stochastic volatility and the Heston model of stochastic volatility, so that

$$d\sigma_t^2 = d \exp(\theta Z_t), \tag{2.17}$$

$$d\sigma_t^2 = \lambda(\omega - \sigma_t^2)dt + v\sigma_t dZ_t, \tag{2.18}$$

where W and Z are Brownian motions. Models with stochastic volatility driven by the exponential of a nonstationary or nearly-nonstationary process, as in (2.17), have long been a feature of the finance literature (e.g., Hull and White (1987)). On the other hand, Heston (1993) proposed modeling volatility as a Ornstein-Uhlenbeck process, as in (2.18) and is currently widely used in the literature. For this final

volatility process, the parameter values are set as follows. We let

$$\text{cov} \begin{pmatrix} V \\ W \\ Z \end{pmatrix} = \begin{pmatrix} 1 & -0.98 & 0.539 \\ -0.98 & 1 & -0.55 \\ 0.539 & -0.55 & 1 \end{pmatrix}. \quad (2.19)$$

For σ_t^2 , we set $\lambda = 1, \omega = 0.04$ and $v = 0.4$ in volatility model (a), and $\theta = 1/\sqrt{T}$ for volatility model (b). These parameter values are the approximate average parameter estimates given by Bakshi, Cao, and Chen (1997) which compares various methods of estimation.

Table III shows the effects of a single structural break in volatility. The table gives the size of the standard OLS t -test, Cauchy t -test, the Bonferroni Q -test, and the REML test at a 5% nominal size when performed on standard calendar time data. Since, as demonstrated previously, the distribution of the existing test statistics depends critically on the structure of volatility, we would expect such a structural break to be distortionary. OLS, which is already inappropriate due to the persistence in the regressor and the correlation in innovations, becomes further distorted. In the most extreme case in which the predictor is nonstationary and there is a large break, most tests will reject a true null. For a relatively small break, the Bonferroni Q -test is over sized for small sample sizes and undersized for larger sample sizes. This result is similar to those in the simple, stationary variance case. However, the distortion increases with the size of the break. If $c = 0$ with 5 years of observations and there is a large change in volatility, the 5% test is 150% too large. The REML test seems to suffer more drastic distortions. It is more heavily impacted by a larger break in the standard deviations of volatility. The size of the REML test is also less sensitive to the sample size in that the size is affected only by the size of the break and the degree of nonstationarity. It is interesting that the size distortion decreases and then

increases as the process goes from nonstationary to more and more stationary. On the other hand, the Cauchy t -test performs quite well despite the break.

Though, existing tests performed poorly when there is a break in volatility, we find quite a different result when using our new technique. Table IV gives the sizes of the same test performed on the time-changed data at the 5% nominal level. The OLS t -statistic remains distorted. This is due to the fact that, despite the time change, there remains a correlation in innovations and persistence in the regressor. The efficient Bonferroni Q -test and the REML test perform much better. Verifying the theoretical result, the sizes are nearly identical to the sizes generated from data with normal errors with a constant variance. The REML test, as in the standard case, remains slightly too large when the regressor has a unit root. Likewise, the Bonferroni Q -test, again as in the standard case, begins too large and then gradually becomes slightly too small as the sample size increases. The size of the Cauchy t -statistics on the time changed data is very nearly ideal.

Next, we apply the same techniques to stochastic volatility models in calendar time. The results of these experiments is given by Table V. Again, the Cauchy t -ratio is very close to optimally sized. Surprisingly, the size of the Bonferroni Q -test often *improves* versus the standard constant volatility models. This should not be understood as an improvement in the quality of the test. Indeed, this demonstrates the distortion to the distribution of the test statistic. For the regime switching and exponential stochastic volatility models, the REML test also suffers severe distortion. However, for the Heston model, the performance of both the Bonferroni Q -test and the REML tests is very similar to that in the standard case.

Finally, we apply a time change to the data tested in Table V. These results are given in Table VI. The improvements to the calendar time tests is clear. While the Bonferroni Q -test and the REML test are not always close to the nominal size of 5%,

Table III. Sizes in calendar time. This table demonstrates the effects of a structural break in volatility on the OLS and Cauchy t -tests, the Bonferroni Q -test, and the REML test. The nominal level is 5% and all values are given in terms of percentages. The experiment is replicated 10,000 times.

c	T	$\sigma_1 = 1.5, \sigma_2 = 1$				$\sigma_1 = 3, \sigma_2 = 1$				$\sigma_1 = 4.5, \sigma_2 = 1$			
		OLS	Cauchy t	Bonf. Q	REML	OLS	Cauchy t	Bonf. Q	REML	OLS	Cauchy t	Bonf. Q	REML
0	5	37.19	4.96	8.30	9.01	49.75	4.77	11.18	14.07	53.37	4.42	12.46	15.88
	20	36.77	4.86	5.62	9.37	49.72	4.76	7.50	13.00	53.72	4.97	8.43	15.94
	50	36.29	4.75	5.20	8.75	49.58	5.14	6.47	13.18	53.68	5.23	7.19	15.37
-2	5	9.79	5.03	8.00	4.69	13.59	5.04	11.04	5.77	14.81	4.92	12.20	6.82
	20	7.58	4.93	4.95	4.79	12.98	5.00	8.36	6.39	14.27	5.09	9.84	6.93
	50	7.00	5.47	4.88	4.80	12.21	4.74	7.62	6.32	14.11	5.12	9.23	6.45
-20	5	6.71	5.42	9.71	6.11	12.32	5.05	10.26	10.43	13.53	4.95	10.57	12.12
	20	6.68	5.33	4.24	6.31	12.88	4.97	4.80	10.37	14.47	4.85	6.57	11.92
	50	6.67	4.89	3.59	6.34	12.29	5.14	6.22	10.04	15.02	5.09	7.12	11.97

Table IV. Sizes in volatility time. This table demonstrates the detrimental effects of a structural break in volatility on the OLS and Cauchy t -tests, the Bonferroni Q -test, and the REML test when applied to the volatility time data. The nominal level is 5% and all values are given in terms of percentages. The experiment is replicated 10,000 times.

c	T	$\sigma_1 = 1.5, \sigma_2 = 1$				$\sigma_1 = 3, \sigma_2 = 1$				$\sigma_1 = 4.5, \sigma_2 = 1$			
		OLS	Cauchy t	Bonf. Q	REML	OLS	Cauchy t	Bonf. Q	REML	OLS	Cauchy t	Bonf. Q	REML
0	5	30.27	4.88	9.45	8.76	30.75	4.81	9.03	8.02	29.88	5.07	8.86	8.29
	20	29.47	5.13	5.52	8.46	29.53	5.41	5.65	8.31	29.14	4.97	5.80	7.99
	50	29.23	5.27	4.86	8.13	29.63	5.02	4.85	8.69	29.05	4.92	4.52	7.60
-2	5	19.49	5.10	8.83	5.23	21.63	5.05	9.07	5.85	23.45	5.24	9.04	4.34
	20	18.23	4.73	5.39	5.23	20.98	4.95	5.73	5.63	20.00	5.01	5.73	4.60
	50	19.17	4.95	4.69	5.57	20.57	4.79	4.90	5.87	21.32	4.77	4.43	4.83
-20	5	7.54	5.12	11.61	5.19	8.32	4.87	12.64	5.25	8.82	4.93	12.36	4.97
	20	7.55	5.26	4.37	5.30	7.96	5.00	4.03	5.83	8.48	4.98	4.39	5.20
	50	7.48	5.64	3.42	5.38	8.28	4.68	3.76	5.61	8.58	5.03	3.56	4.86

Table V. Sizes in calendar time. This table demonstrates the effects of nonstationary volatility on the OLS and Cauchy t -tests, the Bonferroni Q -test, and the REML test. The nominal level is 5% and all values are given in terms of percentages. The experiment is replicated 10,000 times.

c	T	Regime Switching					Exponential Stochastic Volatility					Heston Model				
		OLS	Cauchy t	Bonf. t	Bonf. Q	REML	OLS	Cauchy t	Bonf. t	Bonf. Q	REML	OLS	Cauchy t	Bonf. t	Bonf. Q	REML
0	5	28.84	5.03	7.39	7.39	7.11	12.43	4.88	6.48	6.48	9.67	6.31	5.21	7.38	7.38	5.75
	20	28.96	4.90	6.50	6.50	7.26	12.32	4.87	6.59	6.59	9.70	5.41	4.97	6.49	6.49	4.89
	50	29.03	5.04	6.10	6.10	6.36	12.32	5.01	7.05	7.05	9.49	5.64	5.27	5.91	5.91	5.40
-2	5	30.43	4.94	7.89	7.89	7.15	12.67	5.02	6.25	6.25	9.43	6.33	5.07	7.29	7.29	5.56
	20	28.78	4.54	7.51	7.51	7.78	11.88	4.90	6.26	6.26	9.02	5.21	4.81	5.86	5.86	4.90
	50	28.45	5.09	6.67	6.67	7.18	11.92	4.63	6.56	6.56	9.00	5.49	5.16	5.66	5.66	5.26
-20	5	25.95	5.29	9.54	9.54	8.48	11.35	5.10	7.09	7.09	10.25	5.89	5.06	8.11	8.11	5.27
	20	23.89	5.04	9.90	9.90	11.81	10.85	5.05	7.08	7.08	10.52	5.23	5.30	5.73	5.73	5.11
	50	24.46	5.49	8.97	8.97	11.54	10.62	5.23	7.53	7.53	10.53	4.90	4.85	5.90	5.90	4.81

they are indistinguishable from their sizes in the standard case they were designed to address. That is, they perform as designed since the data now meets their basic assumptions. The Cauchy t -test keeps its outstanding performance in terms of size. Of course, as stated before, this comes at the cost of some power.

As can be clearly seen, when applied to volatility time data, existing tests perform as designed. Conversely, if the data is not well behaved, their performance quickly deteriorates. The existing more efficient tests do have a power advantage over the Cauchy t -test. Since volatility time returns have a normal distribution, we would expect that their power differentials should remain unchanged. Indeed, this is precisely what we find in that the powers are virtually identical to Figure 2.

F. Empirical Results

In our model, we let dY_t be the instantaneous excess return at time t . That is, the (risky) stock return less the (riskless) short-term interest rate. Moreover, we let X_t be the value of the earnings-price ratio, dividend-price ratio, short-term interest rate, or some other suspected predictive variable available at time t .

The data set used in our application consists of

- Stock returns from CRSP: NYSE/AMEX value-weighted index over the period 1926/12/01 - 2002/12/31, daily, monthly, quarterly and annual frequencies.
- Risk free rates from CRSP: One month and three month treasury bond rates.
- dividend-price ratios from CRSP: Dividends over the past year divided by the current price.
- earnings-price ratios from Global Financial Data : A moving average of earnings over the past ten years divided by the current price.

Table VI. Sizes in volatility time. This table demonstrates the effects of nonstationary volatility on the OLS and Cauchy t -tests, the Bonferroni Q -test, and the REML test when applied to the time changed data. The nominal level is 5% and all values are given in terms of percentages. The experiment is replicated 10,000 times.

c	T	Regime Switching				Exponential Stochastic Volatility				Heston Model			
		OLS	Cauchy t	Bonf. Q	REML	OLS	Cauchy t	Bonf. Q	REML	OLS	Cauchy t	Bonf. Q	REML
0	5	29.72	5.06	9.87	8.47	17.30	4.83	9.49	7.43	10.47	4.96	9.16	8.73
	20	29.02	5.04	5.39	8.04	17.26	4.66	5.64	7.50	9.62	5.06	5.71	9.12
	50	29.55	5.25	4.79	8.75	16.79	4.70	4.78	7.49	9.59	5.23	5.03	9.42
-2	5	19.16	5.38	9.20	5.47	14.77	4.67	9.40	7.09	8.92	5.07	9.07	7.30
	20	18.89	5.27	5.72	5.42	14.27	4.70	5.40	7.28	8.22	4.68	5.70	7.75
	50	19.27	4.66	4.47	5.82	13.87	4.76	4.69	6.76	7.60	5.07	4.85	7.42
-20	5	7.77	5.37	11.81	5.38	7.80	4.98	12.11	6.22	6.12	5.33	11.62	5.42
	20	8.10	5.04	4.33	5.49	7.54	5.07	4.34	6.19	5.83	4.88	4.13	5.43
	50	7.80	4.02	3.84	5.56	7.11	4.78	3.81	6.10	5.60	5.10	3.42	5.44

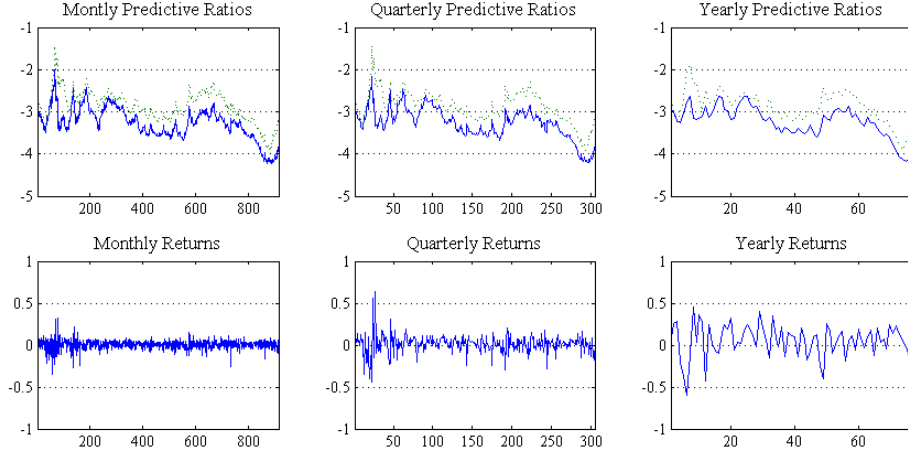


Fig. 3. Data plot. This figure plots the data, with monthly, quarterly, and yearly data going from left to right. The regressors are displayed on the top row and the stock returns are displayed on the bottom. For the regressors, the solid blue line represents the dividend-price ratio, while the dotted green line represents the earnings price ratio.

We investigate the return predictability at monthly, quarterly and annual frequencies. Daily returns adjusted by one month treasury bond rates are used to compute the realized variance of errors. Plots of the data are provided in Figure 3.

For a time change to be appropriate, we must have high frequency data. Daily price data is available, but CRSP dividend-price ratio data is only available at a monthly frequency. To construct a daily dividend-price ratio for day t , we divide the most recent monthly dividend data by the daily price at day t . To provide results corresponding to those reported by Campbell and Yogo (2006), we select Δ so that the number of post-time change observations n are equal to the number of observations in the appropriate Campbell and Yogo (2006) data set. That is, we compare standard time monthly data (for instance) with random time data with volatility equal to the average monthly volatility as measured by realized variance.

We make one practical modification to a time change based strictly on the DDS

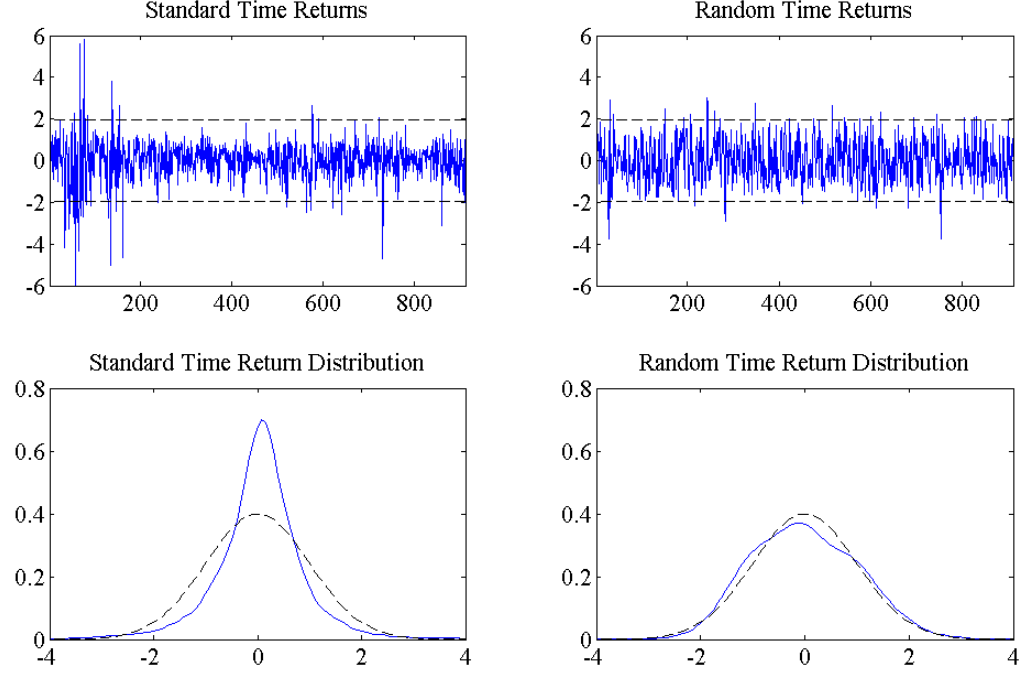


Fig. 4. Data plot and kernel estimates of return distributions before and after the time change. The dashed line represents a standard normal distribution and the blue line gives the estimated distributions for innovations normalized by their respective standard deviations.

definition. As described earlier, the limiting variance of the random time data is exactly Δ . However, if we assign the stopping times based on $\langle Y \rangle_t^\delta > t$, then the actual variance will always be above Δ . This reality is addressed by instead selecting Δ such that we are as close to the limiting variance as possible. That is, we select the stopping times so as to minimize the distance between the realized variance and Δ . Of course, we add the restriction that the time may not repeat; it must go up by at least one period. This technique has been used and detailed previously in Jacewitz, Kim, and Park (2008).

Figure 4 presents the estimated error distributions for the samples collected at

a fixed frequency and under the time change. Recall that we expect the time deformed return process to be normally distributed. The solid lines show the kernel density estimate of the distribution of returns, while the dotted lines in the graphs show the standard normal density. The kernel density estimates are retrieved from the `ksdensity` function in MATLAB. As can be clearly seen in, the error distributions change drastically after the time change. The error distribution from the time-changed regression becomes close to normal, while the distribution of errors from the conventional regression are far from being normally distributed. For the time-changed data, the Kolmogorov-Smirnov statistic is 0.0257 implying a probability of 0.5797, so normality cannot be rejected.

As an additional advantage to the time change, innovations in the regressor become much more normal as well as returns. Normality in innovations is another common and simplifying assumption made in econometric modeling. The normality of innovations when the time change is based on returns is due to a common volatility factor in returns and dividend- or earnings-price ratio. This is obvious since price appears in both the returns and the price ratios. Therefore, since dividends and earnings change relatively little, volatility in returns implies volatility in the predictive ratio. To illustrate this fact, we select parameters μ and ρ which makes $x_{t+1} - \rho x_t$ as close as possible, in the Cramér-von Mises sense, to $N(\mu, \sigma)$, for any μ and σ . We require parameters to be “reasonable”, that is if we impose that $\mu \in [-1, 1]$, and $\rho \in [0.5, 1.1]$. Figure 5 provides the graphical results for the dividend-price ratio. Numerical results are given in Table VII. These show that regardless of the parameter values, the most normal standard time innovations can possibly be is still less normal than the random time innovations. In most cases, the random time distributions are closer to normal than the best possible standard time distributions. The innovations of the earnings-price ratio is marginally closer to normal than for random time for

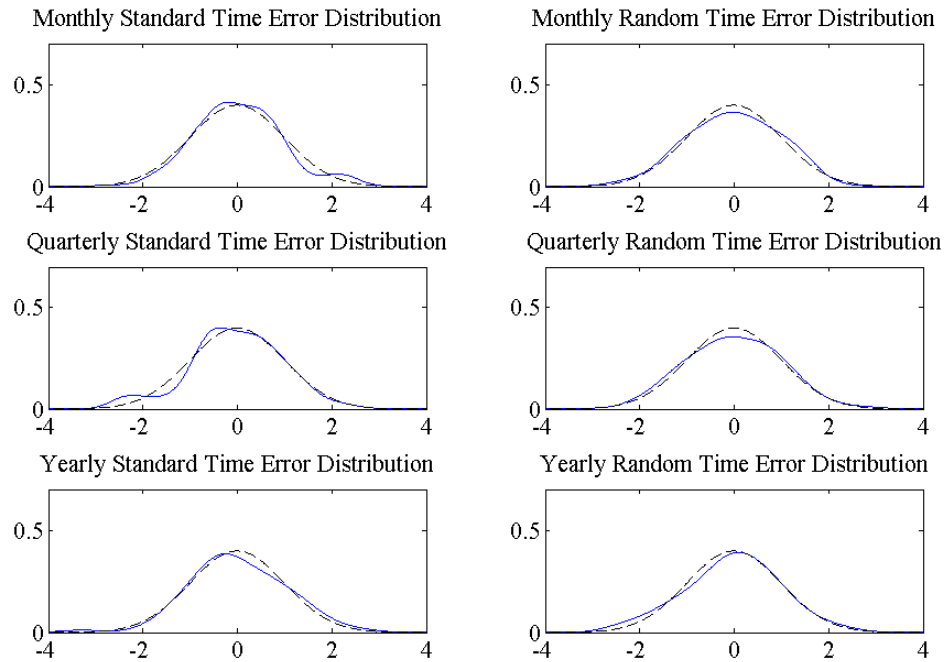


Fig. 5. Kernel estimates of innovation distributions for the parameter values which achieve the most normal innovations in the regressor under standard time sampling and under random time sampling. These are distributions for the dividend-price ratio. For the earnings-price ratio, the results are entirely similar. The dashed line represents a standard normal distribution and the blue line gives the estimated distributions for innovations normalized by their respective standard deviations.

the annual frequency.

Our main results are given in Table VIII. For comparison, the results by Campbell and Yogo (2006) are replicated and reported here. We also add the results for the lagged short-term interest rate, another commonly used regressor. Overall, our estimates of the slope parameter are roughly the same in terms of magnitude as those in previous papers, although ours are more often negative. The important difference is found in the t -statistics. We find that absolutely none of the estimated parameters

Table VII. Results of selecting parameter values which achieve the highest degree of normality. “CvM” refers to the Cramér-von Mises distance from a standard normal distribution.

Frequencies	Covariate	Standard Time			Random Time		
		μ	ρ	CvM	μ	ρ	CvM
Monthly	D/P	0.3265	1.1000	0.9155	-0.3191	1.071	0.0316
	E/P	-0.6586	0.7711	0.4813	-0.3187	1.0707	0.0325
Quarterly	D/P	-1000	0.6949	0.1588	-0.2115	1.0807	0.0330
	E/P	0.2832	1.1000	0.9250	-0.2268	1.0862	0.0327
Annual	D/P	-0.0529	0.9889	0.0378	0.0814	0.9555	0.0311
	E/P	-0.5455	0.8166	0.03037	0.0822	0.9556	0.0306

are significant. Once the time change has been applied, the Cauchy estimates clearly do not support predictability. Moreover, Table IX provides similar results for the subsample periods matching those in Campbell and Yogo (2006). Again, there is no evidence whatsoever for predictability in any subsample.

As previously discussed, the asymptotic theory requires a continuous sample path in the stock price process. This explicitly precludes jumps and other discontinuities. We test for the presence of jumps using bipower variation as described in Barndorff-Nielsen and Shephard (2006) rejecting as discontinuous any interval which displays a test statistic that exceeds the 10% significance level. The offending interval is then eliminated from our analysis. The regression is then run on the remaining data. The results of these procedures are given in Table X. The linear bipower variation test detected 37 and 6 jumps and the ratio test detected 6 and 1 jumps for each interval, respectively. At the 5% level, the linear bipower variation test detected 24

Table VIII. Main empirical results for return predictability. D/P and E/P represents the dividend- and earnings-price ratios, respectively. TC indicates a time change was employed.

Frequencies	Covariate	Campbell and Yogo		TC Cauchy Regression	
		$\hat{\beta}_N$	$\tau(\hat{\beta}_N)$	$\tilde{\beta}_N^{*\delta}$	$\tau(\tilde{\beta}_N^{*\delta})$
Monthly	D/P	0.009	1.71	0.020853	0.046309
	E/P	0.014	2.66	0.55329	1.182
Quarterly	D/P	0.034	2.06	-0.73255	-0.51951
	E/P	0.049	2.91	2.182	1.4949
Annual	D/P	0.125	2.53	-3.3871	-0.57792
	E/P	0.169	2.77	2.3213	0.38705

Table IX. Subsample empirical results for return predictability.

Subsample Range		1926 - 1994		1952 - 2002	
Frequencies	Covariate	$\tilde{\beta}_N^{*\delta}$	$\tau(\tilde{\beta}_N^{*\delta})$	$\tilde{\beta}_N^{*\delta}$	$\tau(\tilde{\beta}_N^{*\delta})$
Monthly	D/P	-0.016721	-0.02899	-0.025263	-0.03695
	E/P	0.72858	1.2156	1.1314	1.5878
Quarterly	D/P	-0.81486	-0.44024	0.44003	0.19657
	E/P	2.908	1.5105	3.8907	1.6675
Annual	D/P	-3.0416	-0.37796	10.4761	0.97535
	E/P	4.532	0.53225	17.702	1.6269

and 5 jumps and the ratio test detected 3 and 0 jumps for each interval, respectively. Finally, at the 1% level, the linear bipower variation test detected 17 and 2 jumps and the ratio test detected no jumps for either interval, respectively. There were no jumps detected at yearly intervals. Still, we find no significant changes in our empirical results. They remain robust to the presence of jumps in the price process.

G. Conclusion

Stock return predictability is one of the most prolific topics in financial economics, and has been for decades. There are some widely recognized features of return and predictive ratio data that can seriously distort standard hypothesis testing. Two characteristics, persistence in the regressor and a correlation between regressand and innovations in the regressor, have been extensively addressed, though no solution has been widely accepted. A third widely accepted characteristic, time varying stochastic volatility, has so far been ignored in the predictive regression literature. These three characteristics can explain the ubiquitous finding of stock return predictability. Our main contribution is to provide a specific technique that is uniquely suited to each of these issues. In the preceding sections, we have offered a new way to test for predictability using a time change to volatility time and the Cauchy estimator.

This new technique has the highly desirable characteristic that, regardless of any of the econometric complications commonly found in the related data, the t -ratio will always have a standard normal limiting distribution. The simple technique consists of constructing a volatility time using realized volatility and then estimating the model using the Cauchy estimator. The Cauchy estimator itself has many useful characteristics provided that the error is independent and normally distributed. The random time sampling ensures that this will always be true.

Table X. Jump robustness check. This table provides estimates of the predictability parameter when intervals which have test statistics below the 10% critical value are removed.

Frequencies	Covariate	Linear		Ratio	
		$\tilde{\beta}_N^{*\delta}$	$\tau(\tilde{\beta}_N^{*\delta})$	$\tilde{\beta}_N^{*\delta}$	$\tau(\tilde{\beta}_N^{*\delta})$
10%					
Monthly	D/P	-0.0078939	-0.57068	-0.0083818	-0.64655
	E/P	-0.0061447	-0.48864	-0.0058486	-0.50149
Quarterly	D/P	-0.021698	-0.28862	-0.021947	-0.29636
	E/P	-0.019398	-0.2813	-0.01729	-0.25649
5%					
Monthly	D/P	-0.0082209	-0.60318	-0.0082877	-0.6411
	E/P	-0.0061745	-0.50529	-0.0062952	-0.54155
Quarterly	D/P	-0.020763	-0.27715	-	-
	E/P	-0.018859	-0.27445	-	-
1%					
Monthly	D/P	-0.0080578	-0.60092	-	-
	E/P	-0.0060236	-0.49953	-	-
Quarterly	D/P	-0.022181	-0.29786	-	-
	E/P	-0.017502	-0.25805	-	-

We demonstrated that these widely hypothesized data characteristics can cause severe over rejection of a true null. With Monte Carlo simulations, we quantified the distortionary effects of stochastic volatility. To empirically apply our technique, we used data covering the same time periods and the same regressors as Campbell and Yogo (2006), who find evidence of predictability. However, we found no evidence, whatsoever, of predictability in stock returns using the most commonly used predictive ratios. Theoretically, this results is unsurprising. Empirically, it is striking. The vast majority of previous papers find strong evidence in favor of predictability, to the extent that it has become a stylized fact of stock returns. In our results, there is no support for predictability at any frequency for any of the predictors examined. The results are strong and unambiguous. It seems clear that the return predictability disappears, if the characteristics of the data are properly addressed.

CHAPTER III

TESTING FOR NO ARBITRAGE IN CONTINUOUS TIME:

A RESOLUTION TO THE FORWARD PREMIUM ANOMALY

We investigate the forward premium anomaly in a continuous time framework. A new model is introduced and then empirically tested using an innovative technique. Given the ultra high-frequency nature of the currency market, a continuous time approach is, in many ways, attractive. We derive a no arbitrage condition for a two country currency exchange market. In particular, the condition implies that the volatility of spot exchange returns, as well as the risk premium requested by spot traders, are functions of differences in the market prices of risk between the two countries. Moreover, it reveals that, even in the simplest case where there is no market price of risk differential, the conventional tests based on discrete time models will be invalid and subject to data misaggregation bias unless the expectation hypothesis holds continuously at all frequencies and maturities.

To empirically evaluate our continuous time regression model, we employ a novel econometric methodology based on a time change from calendar to volatility time. Specifically, our method requires collection of samples at random intervals having the same level of excess exchange return volatility. This amounts to using a sampling chronometer that runs at a rate inversely proportional to the volatility. By doing so, we may effectively make the distributions of residuals into independent standard normals, nonparametrically correcting for the non-normality and time-varying stochastic volatility typically present in exchange return data. The model is estimated by minimizing the distance between the sampling distributions of residuals and the standard normal distribution.

A. Introduction

The forward premium anomaly (FPA hereafter) is widely considered one of the most enduring and significant unsolved puzzles in international finance. Moreover, testing this statement is equivalent to testing the arguably more important uncovered interest parity hypothesis; a condition which is explicitly assumed to be true in many international macroeconomic models, but often fails empirically. Understandably, the relationship between expected and actual future exchange rates interests firms and investors, as well as economists. The forward rate unbiasedness hypothesis states that expected speculative return in the forward exchange market is zero. That is, the forward exchange rate is an unbiased predictor of the future spot exchange rate. The anomaly lies in the ubiquitous conclusion that, while the forward-spot spread is a predictor of the future spot return, it is a reverse predictor. The forward premium mispredicts both the magnitude and, more troublesomely, the sign of future returns.

In light of the ultra-high frequency nature of the currency exchange market, we present a new continuous time forward exchange rate model. Through this specification, we can see directly the effect market prices of risk have on the volatility of exchange rate returns. In addition to deriving the model, we estimate it using the Martingale Estimation Technique recently developed by Park (2008). Once we apply the new econometric technique to the new economic model, we find no evidence for the anomaly. As should be expected, the forward premium is an unbiased predictor of future exchange rate return.

The anomaly originated more than two decades ago (see Longworth (1981)), but remains a dependably prolific area (see Baillie and Kiliç (2006), Mark and Moh (2007), Wu (2007), and Kellard and Sarantis (2008)). The nature of the currency market is among the most important reasons for the popularity and longevity of this

puzzle. The foreign exchange market is extremely active, even by financial market standards. The New York Federal Reserve Bank’s Foreign Exchange Committee (Foreign Exchange Committee 2008) reports that the average daily volume in over-the-counter foreign exchange instruments in North America alone was \$713 billion in the year proceeding April 2007. This includes \$393 billion in spot transactions and \$93 billion in forward transactions. For a reference point for the magnitude of this market, the average daily US GDP in 2007 was around \$38 billion. Globally, the Bank for International Settlements (2007) found that average daily total foreign exchange market turnover \$3.2 trillion. Including \$1 trillion in spot transactions and \$362 billion in forward transactions. Further, the market is highly concentrated. Nearly 86% of all transactions involved the top seven currencies (viz. the currencies which we analyze here: US dollar, euro, yen, pound sterling, Swiss franc, Australian dollar, and Canadian dollar; the so called “majors”). Bjønnes and Rime (2005) list the foreign exchange market as “by far” the largest financial market in existence.

In this paper, we study the FPA in a continuous time framework. In particular, we present a continuous-time version of the FPA model, which is derived using the assumption of no arbitrage from the stochastic differential equation linking spot exchange rate changes to forward spot spread at an infinitesimal time interval. We believe that the no arbitrage condition in continuous time used to derive our model is much more realistic and relevant than the analogous condition in discrete time, given the ultra high-frequency nature of transactions in the currency market. Our model implies, among other things, that both the risk premium and the volatility of spot exchange rates are functions of the difference in the market prices of risk between the two countries. If the two countries have the same market prices of risk, then our model reduces to a continuous time version of the traditional FPA models presuming risk neutrality. However, even in this case, we make it clear that the conventional

FPA tests based on discrete time models will be invalid and subject to data misaggregation bias unless the expectation hypothesis holds continuously at all frequencies and maturities.

Despite the scope of previous literature, our approach has several trenchant differences from some of the most recent contributions in the test of FPA. Mark and Moh (2007) develop a continuous time model of an exchange market. However, we derive a different uncovered interest parity condition (UIP hereafter) derived directly from a no arbitrage condition. Thus, the resulting model is quite different. They relate their model to the FPA via interest rates, while we test the FPA directly by connecting the exchange rate return with the instantaneous forward premium. Both Chaboud and Wright (2005) and Bernoth, von Hagen, and de Vries (2007) study questions closely related to the FPA using a discrete-time UIP condition. Both papers conclude that the UIP tends to hold over very small maturities. Clearly, this result will be closely related with our continuous time setup. It has been difficult to explain why the forward unbiasedness holds at short maturities, but not at long maturities. A continuous time setup sheds some light on this discrepancy. We address this more fully in Section B. Baillie and Kiliç (2006) consider whether nonlinearities can explain the FPA. They form risk adjusted forward premia for risk simply by dividing the standard premia by the standard error. However, this and virtually all previous work does not address the stochastically volatile nature of the data.

Besides introducing a new way to examine the FPA, we use a novel econometric methodology to statistically analyze the model. This methodology relies on random sampling using a time change from calender to volatility time. Our sampling chronometer runs at a rate inversely proportional to the volatility. More precisely, the samples are collected in our analysis at random intervals having the same level of excess exchange return volatility, rather than at some fixed frequency such as daily,

weekly, or monthly. Under this scheme, the samples may be regarded as being independent and identically normally distributed. This is due to the celebrated theorem by Dambis, Dubins and Schwarz, which is well known in the theory of stochastic processes. By using a time change, we may therefore accommodate the non-normality and time-varying stochastic volatility that is typically present in exchange return data. After the time change, our model is estimated and tested by the so-called martingale method developed recently by Park (2008), which minimizes the distance between the sampling distributions of residuals and the standard normal distribution.

The rest of the paper is organized as follows. In the remainder of this section, we will introduce the conventional theory and the anomalous empirical findings reported in previous studies of the FPA. In Section B, we present a new continuous time version of the forward premium equation. The equation is obtained from a continuous time model of a foreign currency market derived under the no arbitrage condition. Section 3 introduces the econometric methodology used in our analysis. We provide a brief overview of the time change and martingale estimation method, which is followed by a detailed explanation of how we may implement them to estimate and test our model. Section 4 reports all of the results for our statistical analysis. Some concluding remarks are given in Section 5. The appendix contains some additional information on our statistical analysis. Finally, a word on notation. In the paper, we deal with both continuous time processes and discrete time series. The time index will be denoted by t or s for the former, and by i and j for the latter throughout the paper.

1. Theory

Let $F_{i,j}$ be the forward exchange rate, that is the time i price denominated in the domestic currency of one unit of foreign currency with a future value date (date of exchange) j . Let X_i be the time i price denominated in the domestic currency of one

unit of foreign currency to be delivered immediately, or the spot exchange rate. The standard hypothesis of unbiasedness is

$$F_{i,j} = \mathbb{E}_i [X_j],$$

or equivalently

$$\frac{F_{i,j}}{X_i} = \frac{\mathbb{E}_i [X_j]}{X_i},$$

for all i and j , where \mathbb{E}_i is the mathematical expectation operator conditional on the information set available at time i . The second equation is the most commonly tested implication of the hypothesis in recent literature.

Let x_i be the log of the spot exchange rate at the i -th period, and $f_{i,j}$ be the log of the forward exchange rate at the i -th period with value date j periods ahead. Unbiasedness of the forward exchange rate would, of course, mean approximately that $\mathbb{E}_i[x_j] = f_{i,j}$. This implies that $f_{i,j} = x_j + u_j$, where u_i is a martingale difference sequence and can be interpreted as the rational agent's prediction error. The equality should hold if agents are risk neutral and rationally use information, there are no transaction costs, and the market is competitive and efficient.

The early works examining forward rate unbiasedness focused on conventional OLS regressions of the most direct test of the unbiasedness hypothesis equation in levels

$$x_j = \alpha + \beta f_{i,j} + u_j. \tag{3.1}$$

Here, the null would be that $\alpha = 0$ and $\beta = 1$. The common finding was that β was, indeed, very close to 1 and the null of unbiasedness could not be rejected. Precisely speaking, however, this should not be regarded as evidence for an unbiased forward price, since both the spot and forward exchange rates are nonstationary. If they are integrated processes, their relationship (3.1) in levels defines a cointegrating

regression, for which the usual OLS estimator is consistent even when endogeneity is present and $\mathbb{E}_i[x_j] \neq f_{i,j}$. It is well known that the OLS procedure is super-consistent, though not fully efficient, in this case. See, e.g., Goodhart, McMahon, and Ngama (1997) for a discussion on how to efficiently analyze the regression (3.1) in levels as a cointegrating regression.

Due to these data characteristics, forward rate unbiasedness has more often been tested based on the regression

$$x_j - x_i = \alpha + \beta (f_{i,j} - x_i) + u_j, \quad (3.2)$$

where $f_{i,j} - x_i$ is the forward premium or discount depending on whether this difference is positive or negative, respectively. Throughout this paper, we will follow the literature and abuse diction by simply referring to $f_{i,j} - x_i$ as the forward premium, irrespective of its sign. It is widely accepted that the return is a stationary process and widely assumed that the forward premium is as well, so (3.2) could then be considered a standard stationary regression which can be analyzed by OLS. We might expect $\alpha = 0$ and $\beta = 1$, if preferences are risk neutral. Clearly, under the null, equations (3.1) and (3.2) are equivalent. Thus, the (log of) the forward premium provides an unbiased forecast of the (log of) the future spot exchange rate return.

Equation (3.2) is derived by combining the Covered Interest Parity (CIP) and the Uncovered Interest Parity (UIP) conditions. The CIP, a fundamental assumption in international finance, relates the forward premium to cross-country interest rate differentials according to

$$f_{i,j} - x_i = r_{i,j} - r_{i,j}^*,$$

where $r_{i,j}$ and $r_{i,j}^*$ are the returns on zero-coupon bonds at the i -th period with maturity at some future time j for the domestic and foreign countries, respectively.

This relationship is widely supported empirically. The UIP, instead, connects the interest rate differentials to the spot exchange rate return, and it is given by

$$x_j - x_i = r_{i,j} - r_{i,j}^*$$

using the same notation as above. It is assumed to be true in many macroeconomic models and yet routinely fails when tested empirically.

A stylized fact of major exchange rates is that they very closely mimic simple random walks. In fact, it is difficult to outperform the simple random walk in forecasting exchange rates. This has led to another common and well known hypothesis that exchange rates follow a random walk. If the exchange rates did follow a random walk, then it follows in particular that the best predictor of the future spot rate is the current spot rate, $x_i = \mathbb{E}_i[x_j]$, in the mean squared error sense. In this case, we should find in (3.2) that $\alpha = 0$ and $\beta = 0$. If $\alpha = 0$ and $\beta \neq 0$ in (3.2), then we should have $f_{i,j} = x_i$, in which case β becomes unidentified.

2. The Anomaly

As expected, the vast majority of FPA papers conclude that the forward premium is, indeed, a predictor of future spot returns. However, the anomalous finding in investigating the unbiasedness hypothesis is the consistent result that the forward premium is (counter-intuitively) a reverse predictor. A plethora of empirical studies have not only rejected the null of unbiasedness, but have found significantly negative estimates of the slope parameter. Thus, if these results are correct when the forward rate is below the spot rate, there is always an expected positive return to entering into a forward contract. Baillie and Bollerslev (2000) believe that: “the FPA has become a well established regularity and is generally regarded as being one of the most important unresolved paradoxes in international finance, and occupies a similar

role to that of the equity premium puzzle in financial economics.” Froot (1990) notes that the average value of over 75 published estimates is -0.88 and very few are positive. Some of these estimates can be quite low. Consider for example an estimate of $\beta = -5.644$ for the United States dollar/Dutch guilder exchange rate given by ?. This would imply that the forward premium is, indeed, a valid predictor of the future spot return. If the current forward exchange rate were 1% higher than the current spot exchange rate, we would expect that on average there would be a -5% loss in the corresponding future spot price of the foreign currency.

These results are troubling on an intuitive and a theoretical level. If we accept this empirical regularity, it can be justified only by abandoning the rational expectations assumption or by the existence of a time-varying risk premium. Further, this time-varying risk premium must exhibit more volatility than the expected depreciation. Such a characteristic is difficult to justify economically. Assume that $f_{i,j} - x_i$ and $x_j - x_i$ are jointly stationary and ergodic. If $\hat{\beta}$ is a consistent estimator, then it follows that

$$\hat{\beta} \rightarrow_p \beta = \frac{\text{cov}(f_{i,j} - x_i, x_j - x_i)}{\text{var}(f_{i,j} - x_i)}.$$

However, we have that $\text{cov}(f_{i,j} - x_i, x_j - x_i) = \text{cov}(f_{i,j} - x_i, \mathbb{E}_i[x_j] - x_i)$ and the forecast error $x_j - \mathbb{E}_i[x_j]$ is orthogonal to $f_{i,j} - x_i$, if we maintain the assumption of rational expectations. Moreover, if we let

$$\nu_{i,j} = f_{i,j} - \mathbb{E}_i[x_j],$$

i.e., the (potentially time varying) risk premium, then

$$\text{cov}(f_{i,j} - x_i, \mathbb{E}_i[x_j] - x_i) = \text{var}(f_{i,j} - x_i) - \text{cov}(\mathbb{E}_i[x_j] - x_i, \nu_{i,j}) - \text{var}(\nu_{i,j}).$$

Therefore, we may write

$$\beta = 1 - \beta_\nu$$

with

$$\beta_\nu = \frac{\text{cov}(\mathbb{E}_i[x_j] - x_i, \nu_{i,j}) + \text{var}(\nu_{i,j})}{\text{var}(f_{i,j} - x_i)}.$$

Consequently, β is low if $\text{cov}(\mathbb{E}_i[x_j] - x_i, \nu_{i,j}) + \text{var}(\nu_{i,j})$ is positive and large. Clearly, if market participants are risk neutral, then $\nu_{i,j}$ would be zero and $\beta = 1$.

A highly variable risk premium is difficult to justify with macroeconomic models since consumption and other determinants of the risk premium are usually smooth. Further economic unpleasantness arises from the consistent finding that the implied levels of risk aversion are unreasonably high. For example, Hodrick (1989) estimates the coefficient to be 60.918. Kaminsky and Peruga (1990) estimate it to be 372.37. A reasonable value used in calibration is around 1.5.

Rejecting (or modifying) the assumption that agents have rational expectations is certainly controversial. Moreover, though several popular models can support a positive slope coefficient that is less than one. (i.e., Bekaert (1996)), none that we know of can justify a time-varying risk premium volatile enough to imply a negative slope coefficient. Thus, much work has focused on showing why the econometric techniques themselves are flawed and explaining how the standard regression could give such strange results.

Without question, the FPA has its share of econometric complications. The forward rate is commonly found to be a highly persistent process. Baillie and Bollerslev (1994) suggest that the temporal dependencies exhibited by the forward premium are well described by a fractionally integrated process. Baillie and Bollerslev (2000) argue that the FPA is merely a statistical artifact created by the persistence in the forward premium and the small sample sizes. Thus, the anomalous finding is merely

due to the relatively slow convergence rate of $\hat{\beta}$. Bekaert, Hodrick, and Marshall (1997) attempt to base hypothesis testing on the small-sample distributions rather than the test statistic's asymptotic distributions. Instead of explaining the puzzling conclusions, their results strengthen the evidence against the expectations hypothesis of the term structure of interest rates.

It appears that there are two basic reactions to the FPA: economic and econometric. Those who subscribe to the former assert that the econometric tools and therefore the results are correct, thus the underlying economic theory must be incorrect. Advocates of the latter, however, claim that the standard economic theory is correct and the econometric methods are then invalid. Thus, economists assume that the econometricians are correct and seek to adjust their fundamental models (e.g. agents are not rational), while the econometricians assume that the economists are correct and adjust their techniques. This paper suggests that both the economic and econometric theory require adjustment. Here, we suggest an appropriate model (which has been done before in various ways), we also suggest an econometric technique suitable to our economic framework (which has also previously been done to varying degrees of success), then we tackle this issue via re-estimating the appropriate model using a technique appropriate to a continuous time framework and the data characteristics exhibited by the spot and forward exchange rates (which has not).

B. A Continuous Time Model of the Foreign Currency Market and Forward Exchange Rates

As Hansen and Hodrick (1980) state, “any discussion of the efficiency of a market requires a specification of the preferences (for risk) and information sets of economic agents,... and the costs inherent in transactions.” Given the highly active and com-

petitive nature of foreign exchange markets and the availability of much improved and inexpensive transactions technology, we model our valuation framework in continuous time, with negligible transaction costs, where information flows via a standard Brownian motion. In absence of consensus on risk preference modeling in general equilibrium asset pricing models, we adopt a no arbitrage condition following much of the finance literature (see Duffie (2001) for details). This method has an advantage of imposing a competitive equilibrium restriction and allowing one to model a risk adjustment for asset returns.

Specifically, consider a continuous-time economy that consists of two countries referred to as the domestic country and the foreign country. Assume that there is no arbitrage in international markets and that the currency market clears continuously. Let W be a standard Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and fix the standard filtration (\mathcal{F}_t) generated by W . We define the spot currency exchange rate X_t by the domestic value per unit of the foreign currency at time t . Further, let this price process, X_t , follow an Ito process driven by the Brownian motion W . Assume that there exist locally riskless money-market accounts B and B^* exclusively available in the domestic and foreign currency, respectively, with the following laws of motion

$$dB_t = r_t B_t dt$$

$$dB_t^* = r_t^* B_t^* dt,$$

where we set $B_0 = B_0^* = 1$ and r_t and r_t^* are the instantaneous short-term interest rates for each country at time t .

If we further assume that markets are complete, we are guaranteed the existence of the unique and equivalent martingale measures $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{P}}^*$ for the domestic and

foreign countries, respectively. In what follows, we will denote respectively by $\tilde{\mathbb{E}}_t$ and $\tilde{\mathbb{E}}_t^*$ the \mathcal{F}_t -conditional expectations with respect to $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{P}}^*$. We can now derive a risk-adjusted uncovered interest parity equation.

Theorem 1 *If there exist no arbitrage across countries and markets are complete, spot exchange return and foreign and domestic spot interest rates will be related according to*

$$\frac{X_t}{X_0} = \frac{\exp\left(-\int_0^t r_s^* ds\right) D_t^*}{\exp\left(-\int_0^t r_s ds\right) D_t}, \quad (3.3)$$

where D_t and D_t^* are the Radon-Nikodym derivatives of $\tilde{\mathbb{P}}^*$ and $\tilde{\mathbb{P}}$, respectively, with respect to \mathbb{P} on \mathcal{F}_t , i.e., $D_t = \mathbb{E}_t[d\tilde{\mathbb{P}}/d\mathbb{P}]$ and $D_t^* = \mathbb{E}_t[d\tilde{\mathbb{P}}^*/d\mathbb{P}]$.

Proof of Theorem 1 Suppose that there exists a domestic asset whose current price is S_0 at time 0 and pays S_t at t . Assuming that there is no arbitrage, there exists an equivalent martingale measure $\tilde{\mathbb{P}}$ such that the deflated process is a martingale. That is,

$$S_0 = \tilde{\mathbb{E}}_0 \left[\frac{S_t}{B_t} \right]. \quad (3.4)$$

For the same asset, in the foreign country, we have

$$S_0/X_0 = \tilde{\mathbb{E}}_0^* \left[\frac{S_t/X_t}{B_t^*} \right],$$

i.e.,

$$S_0 = \tilde{\mathbb{E}}_0^* \left[\frac{S_t(X_0/X_t)}{B_t^*} \right] \quad (3.5)$$

under the no arbitrage condition.

It follows from (3.4) and (3.5) that

$$\begin{aligned}\int_A S_0 d\tilde{\mathbb{P}} &= \int_A \frac{S_t}{B_t} d\tilde{\mathbb{P}} \\ \int_A S_0 d\tilde{\mathbb{P}}^* &= \int_A \frac{S_t(X_0/X_t)}{B_t^*} d\tilde{\mathbb{P}}^*\end{aligned}$$

for any $A \in \mathcal{F}_0$. Moreover, we have

$$\int_A S_0 d\tilde{\mathbb{P}}^* = \int_A S_0 \frac{D_0^*}{D_0} d\tilde{\mathbb{P}} = \int_A S_0 d\tilde{\mathbb{P}}$$

and

$$\int_A \frac{S_t(X_0/X_t)}{B_t^*} d\tilde{\mathbb{P}}^* = \int_A \frac{S_t(X_0/X_t)}{B_t^*} \frac{D_t^*}{D_t} d\tilde{\mathbb{P}}.$$

Therefore, we have

$$\int_A \frac{S_t}{B_t} d\tilde{\mathbb{P}} = \int_A \frac{S_t(X_0/X_t)}{B_t^*} \frac{D_t^*}{D_t} d\tilde{\mathbb{P}} \quad (3.6)$$

for all $A \in \mathcal{F}_0$.

To finish the proof, note that (3.6) holds for all \mathcal{F}_t -measurable S_t , since the market is assumed to be complete. Consequently, we may readily deduce that

$$\frac{1}{B_t} = \frac{X_0/X_t}{B_t^*} \frac{D_t^*}{D_t},$$

and the stated result follows immediately. ■

One can easily see that the equation (3.3) becomes the usual uncovered interest parity theorem if we assume risk neutrality. Given the completeness of the market and absence of arbitrage, the market prices of risk λ_t and λ_t^* exist uniquely and we may write the Doléans exponentials

$$D_t = \exp \left(- \int_0^t \lambda_s dW_s - \frac{1}{2} \int_0^t \lambda_s^2 ds \right), \quad (3.7)$$

$$D_t^* = \exp \left(- \int_0^t \lambda_s^* dW_s - \frac{1}{2} \int_0^t \lambda_s^{*2} ds \right), \quad (3.8)$$

due to Girsanov's theorem. It is well known that the Novikov condition suffices for these exponentials to be martingale and have finite variance. Plugging (3.7) and (3.8) into (3.3) and then applying Ito's lemma gives

$$d \ln X_t = \left(r_t - r_t^* - \frac{1}{2}(\lambda_t - \lambda_t^*)(\lambda_t + \lambda_t^*) \right) dt + (\lambda_t - \lambda_t^*) dW_t. \quad (3.9)$$

It is clear from (3.9) that, as opposed to the usual UIP equation, the stochastic differential equation of X_t will also include an additional term to $(r_t - r_t^*)$, which is the risk premium related to investing in foreign currency.¹ Note that the instantaneous volatility term is solely determined by the difference of market prices of risk for the two countries' currencies. The risk premium, together with logarithmic adjustment term, is also expressed entirely in terms of the price of risk in each country. Since the instantaneous volatility term is the difference of the two market prices of risk, modeling the risk premium for holding foreign currency reduces to specifying a market price of risk for either the domestic or foreign country. One caveat of this setup is that our model abstracts away from many issues such as capital control, market integration, and other politico-economic risk factors which can affect exchange rate dynamics in a non-trivial way.

We can now derive a continuous-time analogue of the uncovered interest parity equation via forward spot spread. As before, we define $x_t = \ln X_t$ and $f_{t,s} = \ln F_{t,s}$, the log forward price of the foreign currency for exchange at $s > t$.

Theorem 2 If there exist no arbitrage across countries, the domestic and foreign instantaneous interest rate differential and the forward premium will be related ac-

¹Backus, Foresi, and Telmer (2001) in their discrete-time setup with an affine term structure derived a similar expression. But, we do not rely on any particular term structure model nor specific Ito processes to derive our expression.

cording to

$$r_t - r_t^* = \pi_t, \quad (3.10)$$

where π_t is the instantaneous forward premium at time t defined as

$$\pi_t = \lim_{s \rightarrow t^+} \frac{f_{t,s} - x_t}{s - t}.$$

Furthermore, we have

$$dx_t = [\pi_t + (\lambda_t - \lambda_t^*)^2 / 2 - \lambda_t (\lambda_t - \lambda_t^*)] dt + (\lambda_t - \lambda_t^*) dW_t \quad (3.11)$$

in terms of the instantaneous forward premium and risk premium.

Proof of Theorem 2 If we denote $Z(t, s)$ and $Z^*(t, s)$ as the zero coupon bond prices for the domestic and foreign countries with maturity $s > t$, a simple arbitrage argument implies that

$$\frac{F_{t,s}}{X_t} = \frac{Z^*(t, s)}{Z(t, s)}, \quad (3.12)$$

i.e.,

$$f_{t,s} - x_t = \ln Z^*(t, s) - \ln Z(t, s). \quad (3.13)$$

Recall that

$$r_t = - \lim_{s \rightarrow t^+} \frac{\ln Z(t, s)}{s - t}.$$

Therefore, it follows by definition that

$$\begin{aligned} r_t - r_t^* &= \lim_{s \rightarrow t^+} \frac{\ln Z^*(t, s) - \ln Z(t, s)}{s - t} \\ &= \lim_{s \rightarrow t^+} \frac{f_{t,s} - x_t}{s - t}. \end{aligned} \quad (3.14)$$

If we combine (3.9) and (3.10), we have our equation of exchange rate determination (3.11). ■

Note that π_t , the instantaneous forward premium at time t , represents the premium paid to an agent who agrees to a contract to exchange currencies at a certain price an infinitesimally small time in the future. Therefore, (3.14) is simply an instantaneous version of the well known CIP equation. That is, for an arbitrarily small, discrete interval $[t, t + \delta]$, (3.14) is

$$(r_t - r_t^*) \delta \simeq \ln F_{t,t+\delta} - \ln X_t.$$

In this sense, (3.11) can be understood as a continuous-time version of forward premium equation with time-varying risk premium and volatility. As mentioned earlier, our continuous-time framework presumes that currency trading occurs at a very fine time scale and markets clear almost continuously. These presumptions are made wholly credible by the nature of the currency market. With hundreds of dealers all over the world in the USD/EUR currency pair alone, the foreign exchange market is clearly a twenty four hour market. Moreover, quotes are available on a per second basis.

A discrete-time relationship between, say, a forward premium with a one month maturity and a currency exchange rate changes over a month, and can be derived by integrating (3.11). However, we encounter several issues in empirically evaluating the forward premium equation. First, it is subject to several econometric complications which we will explain in next section. However, there is another important economic issue. To illustrate this, suppose that (3.11) holds at every t and we want to analyze a discrete-time relation for one-month forward premium using the data collected at a fixed time of every month. If we denote the time by t_i for month i and let $x_i = \ln X_{t_i}$, then we have

$$x_{i+1} - x_i = \int_{t_i}^{t_{i+1}} \left[\pi_s + \frac{(\lambda_s - \lambda_s^*)^2}{2} - \lambda_s(\lambda_s - \lambda_s^*) \right] ds + \int_{t_i}^{t_{i+1}} (\lambda_s - \lambda_s^*) dW_s \quad (3.15)$$

A usual forward premium equation used for this estimation is

$$x_{i+1} - x_i = \alpha + \beta(f_{i,i+1} - x_i) + u_{i+1}, \quad (3.16)$$

where $f_{i,i+1}$ is the log one month forward rate at month i . Even ignoring the presence of a risk premium and various econometric issues arising from persistence in the regressor and time varying stochastic volatilities in the error, (3.16) is not generally compatible with (3.15), unless some form of the expectations hypothesis holds continuously. The expectations hypothesis implies that $f_{i,i+1}$ should be a summation of the expected instantaneous forward rates over a month interval plus some constant. Thus, testing (3.16) amounts to a joint test of the forward premium anomaly and the expectations hypothesis, whereas (3.15) is designed solely for the forward premium anomaly. This is a subtle, but important difference: In order to tackle the forward premium equation in pure form, our example suggests that we have to use a theoretical relationship that is compatible with market clearing intervals.

Of course, if market prices of risk are equal for both countries, and there are no additional sources of shocks affecting currency markets, then the expectations hypothesis will hold in each country and therefore this is no concern. But even in this case there still does exist an econometric problem, i.e. need for bias correction due to temporal dependence. However, given ample evidence against the expectations hypothesis (See Campbell and Shiller (1991) for example), and time-varying nature of volatilities, it is more appropriate to use (3.15) when testing the forward premium anomaly.²

²We do not argue that joint testing is unimportant. On the other hand, this is a very fundamental problem that requires a careful consideration. To this end, however, a full specification of the term structure models for both countries is necessary.

C. Econometric Methodology

1. Econometric Model

The econometric model allows for the possibility of econometric error in the forward premium equation in the form of $\omega_t dV_t$ where V_t is a standard Brownian motion independent of W_t . Consequently, we have

$$\begin{aligned} dx_t &= [\pi_t + (\lambda_t - \lambda_t^*)^2 / 2 - \lambda_t (\lambda_t - \lambda_t^*)] dt + (\lambda_t - \lambda_t^*) dW_t + \omega_t dV_t \\ &= [\pi_t + (\lambda_t - \lambda_t^*)^2 / 2 - \lambda_t (\lambda_t - \lambda_t^*)] dt + \sigma_t dU_t, \end{aligned} \quad (3.17)$$

where

$$\sigma_t^2 = (\lambda_t - \lambda_t^*)^2 + \omega_t^2$$

and U_t is a standard Brownian motion.

The expression (3.17) has several interesting features distinguishing it from its conventional discrete time counterpart. First, the mean of spot returns is a function of each country's market prices of risk. Further, the instantaneous volatility term is solely determined by the difference of market prices of risks for the two countries' currencies. Thus, if agents are risk neutral (i.e., $\lambda_t = \lambda_t^* = 0$) or agents are risk averse such that the market prices of risk are always equal across countries (i.e., $\lambda_t = \lambda_t^* \neq 0$), then (3.17) reduces to an expression similar to the standard forward premium equation. The appropriate continuous-time analogue to the traditional model considers the case where $\lambda_t = \lambda_t^*$. In this case, we are left with the equation

$$dx_t = \pi_t dt + \omega_t dV_t,$$

which may be estimated if properly time-aggregated. However, it seems reasonable that risk attitudes, and therefore the prices of risk, might vary across countries. In

such a case, the standard FPA specification will yield spurious estimations.

Our model (3.17) can be estimated if we specify the market prices of risk for two countries. Of course, correctly specifying and accurately estimating the market price of risk is no trivial task. There is a large and active literature on the estimation of the market price of risk λ_t . Innovation in the estimation of this latent variable is beyond the scope of this paper. Instead, to maintain focus on the FPA, we make several simplifying, but progressively less restrictive, assumptions. Furthermore, we rely on the most fundamental proxy of the market price of risk: the Sharpe Ratio. That is

$$\lambda_t = \frac{\mu_t - r_t^f}{\nu_t} \quad \text{and} \quad \lambda_t^* = \frac{\mu_t^* - r_t^{f*}}{\nu_t^*},$$

where μ_t and μ_t^* are expected returns from the risky assets, r_t^f and r_t^{f*} are the risk-free rates in two countries, and ν_t and ν_t^* are measures of risk in each country's risky asset. Therefore, λ_t and λ_t^* are the excess returns per unit of risk in two countries. We simply use actual returns as a proxy for expected returns (which will be correct on average if agents are rational). For the estimation of our model, we use the S&P 500 for the estimation of μ_t and ν_t , and the three-month treasury bill rate as the risk-free rate r_t^f . We measure the risk at time t by the rolling standard deviation of the returns over the month preceding t .

We model the interaction between the foreign and domestic market prices of risk using several very simple specifications for market prices of risk. For a benchmark model, Model 0 assumes that the market prices of risk are always equal (i.e., $\lambda_t - \lambda_t^* = 0$) or that they have at most a negligible effect on returns (i.e., $\gamma = 0$). This model corresponds to the traditional FPA regression. For a more flexible model, we offer Model 1, which assumes that market prices of risk are constant, but the difference is not equal to zero (i.e., $\lambda_t = \lambda \neq \lambda_t^* = \lambda^*$). In Model 2, the market prices of risk

are variable but their difference is constant (i.e., $\lambda_t - \lambda_t^* = \text{const.} \neq 0$). Since the difference is assumed to be constant, we only estimate the market price of risk for the domestic country. Finally, for Model 3, we allow the both market prices of risk to vary with no restriction on differences.

It is important to note that our model (3.17) allows for the presence of time-varying stochastic volatility in the errors. The instantaneous volatility σ_t is in general stochastic and varies over time. This is very consistent with common characteristics of financial data. This aspect of financial data is both widely accepted in the financial literature and largely ignored by the FPA literature. There are at least two important characteristics that we need to pay particular attentions in dealing with the financial data. First, the distribution of the return process is far from normal. The peakiness and heavy-tails found the return process are ubiquitous throughout financial data. Further, there is time-heterogeneity in the volatility of the return process, another common feature of financial data. See the top panel of Figure 1 for the kernel density estimates of the return distributions and Figure 1 for a demonstration of time-varying volatility.

2. Time Change

To effectively deal with the time-varying stochastic volatility in the error process of (3.17), we use a time change. The idea is based on the widely known theorem by Dambis (1965), Dubins and Schwarz (1965) (DDS hereafter). To state and interpret the DDS theorem more precisely, we need to introduce some additional concepts and notations. For a continuous martingale M_t , we define a time change T_t by

$$T_t = \inf \{s > 0 \mid \langle M \rangle_s > t\},$$

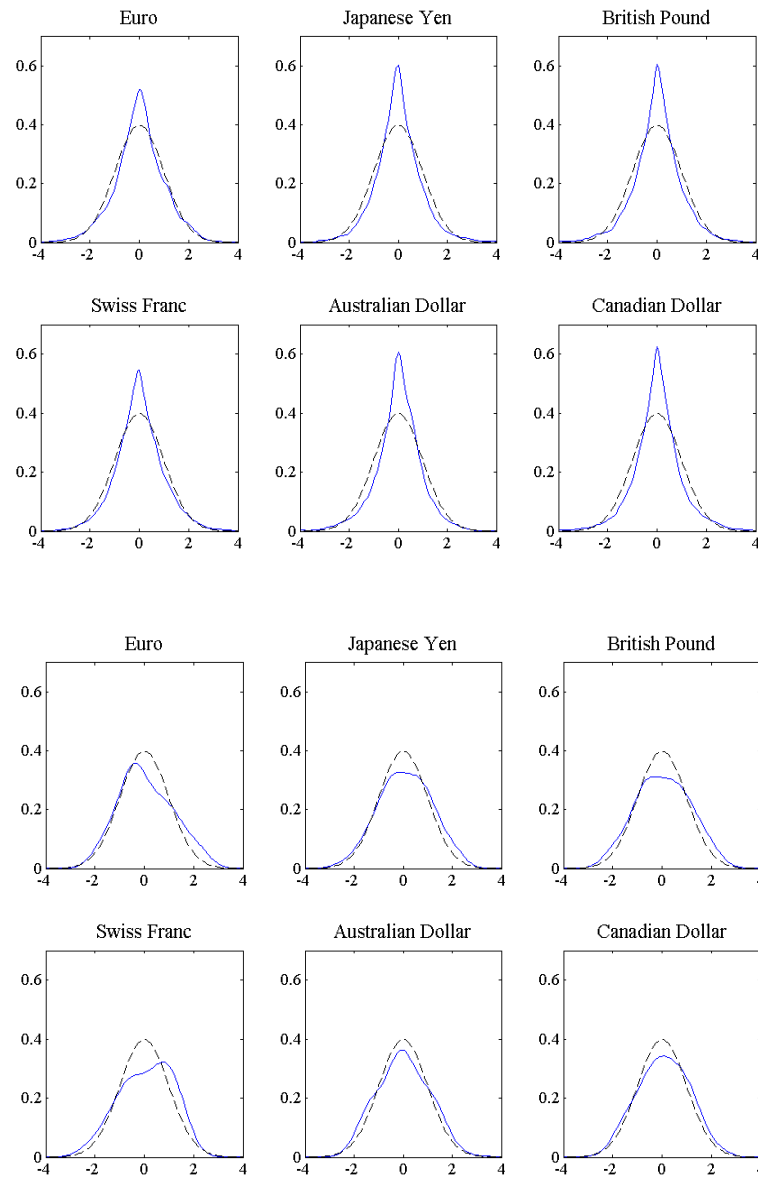


Fig. 6. Kernel density estimates of the return distributions before and after the time change has been performed. The top six panels give the density estimates of the standard time returns for each country. The bottom six panels give the density estimates for the random time returns. The dashed line represents a standard normal distribution.

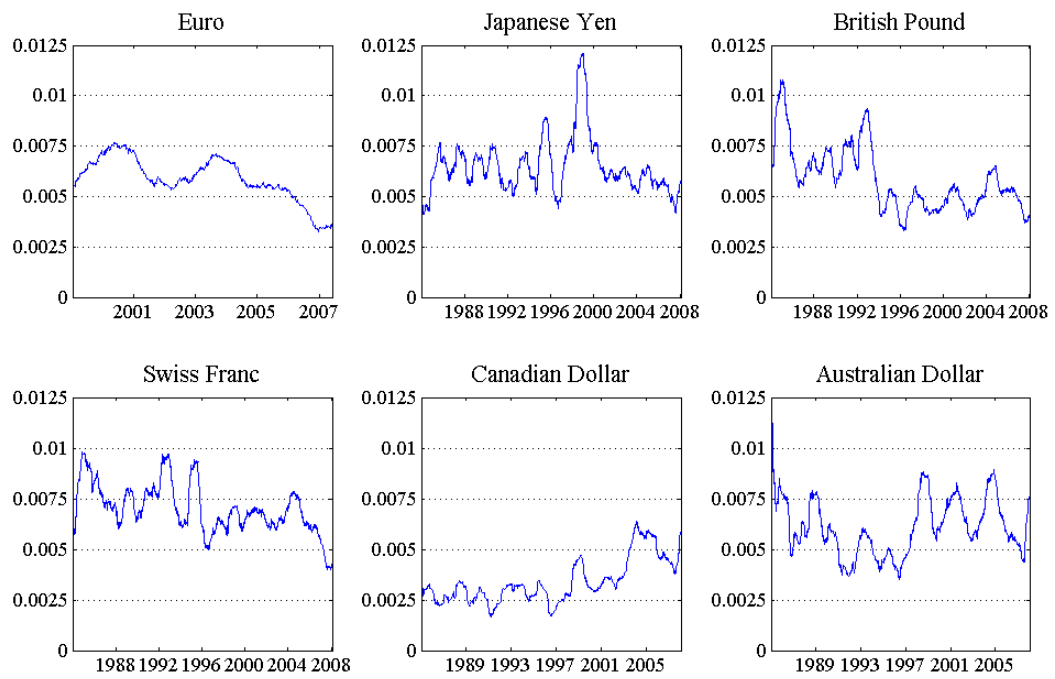


Fig. 7. Daily rolling standard deviation of returns for the previous year.

where $\langle M \rangle_t$ is the quadratic variation process of M_t . If $\langle M \rangle_t$ is continuous and strictly increasing a.s. as in many of the models commonly used for practical applications, the time change T_t is nothing but the time inverse of the quadratic variation process $\langle M \rangle_t$. The DDS theorem states that

$$M_{T_t} = Z_t \quad \text{and} \quad M_t = Z_{\langle M \rangle_t},$$

where Z_t is the standard Brownian motion, which is often referred to as the DDS Brownian motion.

Loosely put, the DDS theorem implies that all continuous martingales are essentially Brownian motions if we use chronometers given by their quadratic variation processes. The standard Brownian motion has a quadratic variation process which is deterministic and given exactly by the actual time. Other more general continuous martingales have quadratic variation processes that are stochastic and varying across their different realizations. Indeed, it is well known that the standard Brownian motion is the only continuous martingale whose quadratic variation process is given by the actual time. It follows from the DDS theorem that if we use a chronometer inversely proportional to its quadratic variation process, any continuous martingale reduces to the standard Brownian motion, or equivalently, that any continuous martingale can be thought of the standard Brownian motion whose sample paths are read using the chronometer given by its quadratic variation process. Thus, we expect that the differences in the process sampled in volatility time, the random time returns, will be normally distributed with a constant variance of our choosing. See the bottom panel of Figure 1 for the kernel density estimates of the random time returns.

We apply the DDS theorem to the continuous martingale process

$$M_t = \int_0^t \sigma_s dU_s,$$

i.e., the error process in our model (3.17), whose quadratic variation process is given by

$$\langle M \rangle_t = \int_0^t \sigma_s^2 ds.$$

Note that we do not impose any restriction on the instantaneous volatility σ_t . In particular, we allow σ_t to be stochastic and varying over time in any arbitrary fashion. Furthermore, it is easy to see that $\langle M \rangle_t$ is continuous and strictly increasing as long as σ_t is non-vanishing except for a set of Lebesgue measure zero.

Now we consider our model (3.17) under the time change, which is given by

$$dx_{T_t} = [\pi_{T_t} + (\lambda - \lambda^*)_{T_t}^2/2 - \lambda(\lambda - \lambda^*)_{T_t}] dT_t + \sigma_{T_t} dU_{T_t} \quad (3.18)$$

and

$$\sigma_{T_t} dU_{T_t} = dM_{T_t} = dZ_t, \quad (3.19)$$

where M_t and Z_t are defined earlier. By employing a time change from calendar time t to volatility time T_t , the error process in our model has become the standard Brownian motion. Therefore, the non-normality and time-varying stochastic volatility in the errors of our original model disappear. Once the clock has changed, the error process has the well behaved features of a constant volatility, independence in increments and Gaussianity. In the next subsection, we will explore this to test for the parameters in the model more effectively.

In general, the error process M_t is not observed. Clearly, it is unobserved in our model (3.17) unless λ_t and λ_t^* are fully specified and observable. The quadratic variation $\langle M \rangle_t$ of the error process, however, is observed without specifying λ_t and λ_t^* . In fact, we have

$$\langle x \rangle_t = \langle M \rangle_t,$$

since the term including π_t , λ_t , and λ_t^* is of bounded variation and its quadratic

variation vanishes at all $t > 0$. Therefore, we may obtain the quadratic variation of M_t directly from the log exchange rate process x_t . For any $t > 0$, the quadratic variation $\langle M \rangle_t$ is estimated by the realized variance

$$\sum_{i=1}^n (M_{t_i} - M_{t_{i-1}})^2,$$

where $0 = t_0 < \dots < t_n = t$, and the time change T_t is obtained from the estimated $\langle M \rangle_t$. For each $t > 0$, the realized variance converges in probability to the quadratic variation as $\max_i |t_i - t_{i-1}| \rightarrow 0$. Therefore, the estimated time change is also expected to converge in probability to T_t . The reader is referred to Park (2008) for the technical details. See Figure 2 for an example of the estimated quadratic variation process and time change for the British pound. In the subsequent explanation of our methodology, we simply assume to ease the exposition that $\langle M \rangle_t$ and T_t are directly observed.

3. Martingale Estimation

For a fixed $\Delta > 0$, it follows from (3.18) and (3.19) that

$$x_{T_{i\Delta}} - x_{T_{(i-1)\Delta}} = \int_{T_{(i-1)\Delta}}^{T_{i\Delta}} [\pi_t + (\lambda_t - \lambda_t^*)^2/2 - \lambda_t(\lambda_t - \lambda_t^*)] dt + \varepsilon_i \quad (3.20)$$

where

$$\varepsilon_i = (Z_{T_{i\Delta}} - Z_{T_{(i-1)\Delta}})$$

that are independent and identically distributed as $\mathbb{N}(0, \Delta)$ for $i = 1, \dots, N$.

Consider Model 1 introduced earlier, we consider the regression model given by

$$x_{T_{i\Delta}} - x_{T_{(i-1)\Delta}} = \alpha_0 + \beta_0 \int_{T_{(i-1)\Delta}}^{T_{i\Delta}} \pi_t dt + \gamma_0(T_{i\Delta} - T_{(i-1)\Delta}) + \varepsilon_i, \quad (3.21)$$

where $\alpha_0 = 0$, $\beta_0 = 1$ and

$$\gamma_0 = (\lambda_t - \lambda_t^*)^2/2 - \lambda_t(\lambda_t - \lambda_t^*),$$

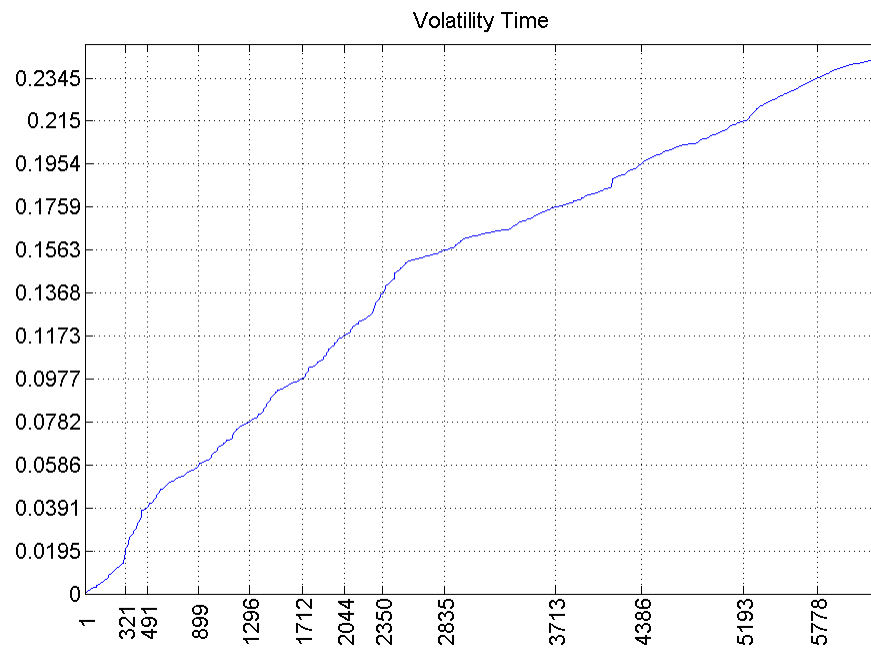


Fig. 8. Calendar versus volatility time. This figure demonstrates a conversion from calendar time to volatility time for the British pound. The vertical axis maps the realized volatility in regular increments. Meanwhile, on the horizontal axis is the time between observations is a random variable.

which is assumed to be a constant for all t . This follows directly from (3.20). Note that the true parameter values α_0, β_0 and γ_0 in (3.21) are identified by the conditions that $(\varepsilon_i/\sqrt{\Delta})$ are independent standard normals. It is clear that the normalized regression error $(\varepsilon_i/\sqrt{\Delta})$ is distributed as independent standard normals for no other values of these parameters.

To estimate the parameter $\theta = (\alpha, \beta, \gamma)' \in \Theta$, we employ the martingale estimation method proposed recently by Park (2008). To introduce the method, we define

$$z_i(\theta) = \frac{1}{\sqrt{\Delta}} \left[x_{T_{i\Delta}} - x_{T_{(i-1)\Delta}} - \alpha - \beta \int_{T_{(i-1)\Delta}}^{T_{i\Delta}} \pi_t dt - \gamma(T_{i\Delta} - T_{(i-1)\Delta}) \right],$$

and let

$$z_i^d(\theta) = (z_i(\theta), z_{i-1}(\theta), \dots, z_{i-d+1}(\theta))',$$

the vector consisting of d -number of consecutive values of $z_i(\theta)$ for each i . Furthermore, we signify by $\Pi_N(\cdot, \theta)$ the empirical distribution function of $(z_i^d(\theta))$ for each $\theta \in \Theta$ and by $\Pi_0(\cdot)$ the distribution function of $(z_i^d(\theta_0))$.

The martingale estimator (MGE) $\hat{\theta}_N$ of the parameter θ is defined as

$$\hat{\theta}_N = \operatorname{argmin}_{\theta \in \Theta} \int_{-\infty}^{\infty} [\Pi_N(z, \theta) - \Pi_0(z)]^2 \varpi(dz),$$

where ϖ is some weight measure, and N is the number of observations selected after the time change. We have that as $N \rightarrow \infty$,

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \rightarrow_d \mathbb{N}(0, \Omega),$$

under suitable regularity conditions, where $\Omega = B^{-1}AB^{-1}$ with

$$A = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dot{\Pi}_0(x) \Sigma(x, y) \dot{\Pi}_0(y)' \varpi(dx) \varpi(dy)$$

$$B = \int_{-\infty}^{\infty} \dot{\Pi}_0(z) \dot{\Pi}_0(z)' \varpi(dz).$$

Here, $\dot{\Pi}_0$ is the derivative of Π_0 , and Σ is the covariance kernel of the limit Gaussian process given in Park (2008).

The motivation for the MGE is surprisingly simple and straightforward. We just find the parameters for which the empirical distribution of the errors get us as close as possible to a standard normal distribution. More precisely, we locate the value of $\theta \in \Theta$ for which $(z_i^d(\theta))$ is closest to the distribution of $(z_i^d(\theta_0))$. Recall that the distribution of $(z_i(\theta))$ becomes independent standard normal only when $\theta = \theta_0$. Therefore, $\theta_0 \in \Theta$ is uniquely given by the distribution of $(z_i^d(\theta_0))$. Of course, the MGE may be viewed as a usual minimum distance estimator, where the distance is given by the Cramér-von Mises (CvM) distance between the empirical distribution of the sample $(z_i^d(\theta))$ with the unknown parameter value $\theta \in \Theta$ and the distribution under the true parameter value $\theta_0 \in \Theta$. In this paper, we use the MGE with $d = 1$.

For the weight measure ϖ , we use the measure given by Π_0 , i.e., $\varpi(dz) = d\Pi_0(z)$. In this case, there are simple ways to compute the CvM distance. To introduce them, we let Φ be the standard normal distribution function, and define

$$t_i(\theta) = \Phi(z_i(\theta)).$$

For $d = 1$, the MGE can be obtained by numerically solving

$$\hat{\theta}_N = \operatorname{argmin}_{\theta \in \Theta} \sum_{i=1}^N \left[t_i(\theta) - \frac{2i-1}{2N} \right]^2$$

For these numerical optimization problem, we set the initial values of parameters α, β

and γ to be given by $\alpha = 0, \beta = 1$, and $\gamma = 0$.

This gives us our estimates of α , β , and γ for Model 1. We may proceed similarly for Models 0, 2, and 3. For Model 2, we consider the regression model in which there may be changes in risk attitudes, but the domestic and foreign market prices of risk stay constant relative to each other. This setup gives us the equation

$$x_{T_{i\Delta}} - x_{T_{(i-1)\Delta}} = \alpha_0 + \beta_0 \int_{T_{(i-1)\Delta}}^{T_{i\Delta}} \pi_t dt + (\gamma_0^2/2)(T_{i\Delta} - T_{(i-1)\Delta}) + \gamma_0 \int_{T_{(i-1)\Delta}}^{T_{i\Delta}} \lambda_t dt + \varepsilon_i, \quad (3.22)$$

where $\alpha_0 = 0, \beta_0 = 1$ and $\gamma_0 = -(\lambda_t - \lambda_t^*)$, which is assumed to be constant. Finally, for Model 3, we allow both prices of risk to vary freely.

$$\begin{aligned} x_{T_{i\Delta}} - x_{T_{(i-1)\Delta}} - \frac{1}{2} \int_{T_{(i-1)\Delta}}^{T_{i\Delta}} (\lambda_t - \lambda_t^*)^2 dt + \int_{T_{(i-1)\Delta}}^{T_{i\Delta}} \lambda_t (\lambda_t - \lambda_t^*) dt \\ = \alpha_0 + \beta_0 \int_{T_{(i-1)\Delta}}^{T_{i\Delta}} \pi_t dt + \varepsilon_i, \end{aligned} \quad (3.23)$$

where $\alpha_0 = 0$ and $\beta_0 = 1$.

D. Empirical Results

Before preceding, let us concretize the various institutional concepts related to the present question. A forward is an contract between parties to exchange currency at a future date (three days or more in the future) at an exchange rate agreed upon today. This three day minimum comes from the practical fact that spot purchases of currency are for delivery in two days. We will refer to the day on which the currency is to be delivered as the value date. A one month forward contract corresponds to 30 days from the current value date, if that day is a business day. If not, it corresponds to the nearest business day that is more than 30 days in the future (see Bekaert and Hodrick (1993) for a detailed description). Forward transactions are often non-standard, however there do exist standard contracts for one month, two months, three

months, six months, and one year which are widely available.

1. Data

Data was gathered from Barclays Bank PLC and retrieved via Datastream. We use daily observations for the US dollar, euro, yen, pound sterling, Swiss franc, Australian dollar, and Canadian dollar. These currency pairs are collectively referred to as “the majors”. Nearly 86% of all currency exchange transactions involve these 7 currencies. The periods are from January 2, 1984 to December 31, 2007, or 6,261 observations for the yen, pound, and franc. The Australian and Canadian dollar data spans January 1, 1985 to December 31, 2007, or 6,000 observations. For modeling purposes, we let the US be the domestic country in all cases.

The euro was introduced on January 1, 1999, giving us less than half of the observations before December 31, 2007 than are available for the other currencies, or 2,345 observations. Estimation of the drift component of any diffusion process relies on the span of the data rather than the frequency. As such, the youth of the euro is an undesirable characteristic. However, the US dollar-euro currency pair is the most traded. We therefore include it in the results for the sake of completeness. We would, of course, expect estimates to be less reliable.

We use the S&P 500 as the domestic risky asset, spanning the data range of the associated foreign index. For Model 3, a foreign market price of risk must also be estimated. For this, we use the FTSE Euro 100, the NIKKEI 225, and the FTSE UK 100 as the risky asset for Europe, Japan, and the United Kingdom. We select these three alone because they have the most developed stock markets with the most available data.

2. Estimation Procedure

In the next subsection, we directly test the forward unbiasedness hypothesis. However, before preceding to the parameter estimates for the actual exchange rate data, let us explain more precisely how the time change and martingale estimate are empirically employed. There are, of course, some additional practical concerns which must be addressed.

The theoretical time-change results hold for any constant value Δ . Of course, in practice some Δ 's are better than others. For instance, a time-change based on a Δ smaller than any of the realized volatilities between observations will yield the original data, which has been shown to be non-normal and to potentially exhibit stochastic volatility. Conversely, if we choose Δ to be the total realized volatility, we are left with one observation. Currently, there is no literature on the optimal value of Δ . We approach this selection in a simple way.

First, we select a Δ based on minimizing the CvM statistic of the returns. The motivation for this being that the time-change of the return process theoretically and ideally will yield a normal distribution. So, we select Δ which gives returns that are closest to that ideal. Further, the Δ must maintain at least 30 post-time-change observations and at least an average of 30 observations per random-time period, or $N = [30, n/30]$. This range is purely arbitrary, but we must ensure that there are a “reasonable” number of data points. The upper limit is to ensure that the average random time interval has enough observations to render estimates of quadratic variation to be reliable to some degree. However, all of the minimized post-time change samples selected are away from these boundaries, which suggests that the constraints are not binding. In practice, it was computationally more expedient to select N rather than directly selecting Δ . Though, these are equivalent since $N\Delta$

is equal to the total realized volatility, which is known and given by the data. The actual selection was made by calculating the CvM statistic or standard errors for every value of N within the range given above.

In the limit, the quadratic variation achieved that surpasses Δ will be infinitesimally close to Δ . In practice, we suspect that assigning the random time based on this rule will lead to an upward bias since the actual realized volatility must always be strictly larger than Δ . To address this issue, we select the time change based on minimizing the distance between the realized volatility and Δ . This way, we get as close to the theoretical limit of quadratic variation within each random time period as possible.

We use subsampling (see Politis and Romano (1993)) to generate standard errors. While we do lose observations in changing chronometers, Brown, DasGupta, Marden, and Politis (2004) have shown subsampling to yield good results even in extremely small samples. We divide our N post-time change samples into overlapping blocks of size b , so that the first subsample has observations 1 through b , the second has observations 2 through $b + 1$, and so forth until the last block with observations $N - b + 1$ through N . We scanned the in the range $[10, N - 10]$ to ensure that there were at least 10 observations per subsample, and that there were at least 10 subsamples. Specifically, we choose the subsample size that gives the smallest variation in estimates. Standard errors are computed by simply using the standard deviation of the estimates of the parameters across all subsamples for each selected subsample size. To be perfectly explicit, the procedure went as follows: (1) we found the number of post-time change observations which gives the lowest CvM statistic, (2) then we found the subsample size b which yields the smallest volatility in estimates of β given N , and finally (3) we computed the parameter estimates using the selected N and computed standard errors using the selected b .

3. Estimates

The parameter estimates for all models were computed by numerically minimizing the CvM distance initialized at the null hypothesized values. The optimization was performed using the `fminsearch` function of MATLAB. For all minimizations, the tolerance was set to 10^{-8} with no maximum number of iterations. For Model 2, we must estimate an additional parameter γ , which corresponds to the domestic constant price of risk. Due to the nonlinear constraint on γ , we used `fmincon` as the optimization routine. The upper and lower bounds were for parameter estimates was 50 and -50. Throughout this section, we will refer to estimates as significant if they are significant at the 5% level.

Table XI shows estimates for our benchmark Model 0. This model assumes that there is no difference in the market prices of risk across countries or that it has negligible effect on returns. Again, this corresponds to the continuous time version of the standard FPA regression. We provide this “naïve” model in order to accentuate the differences between our model and the traditional model. These estimates also show the impact that ignoring the market prices of risk has on parameter estimates, even when using the MGE. The estimates are quite variable; ranging from -0.7077 for the Swiss Franc to 1.4344 for the euro.

Table XII reports our estimations of the simplest new model, Model 1. Firstly, it should be surprising that almost all estimates are positive, and significantly so. This is a unusual finding. Except for the Canadian dollar, all the bias coefficients are significantly different from zero. The pound, Swiss franc, and the Canadian dollar are significantly different from the null of $\beta = 1$. However, it should be noted that this basic model makes the arguably unreasonable assumption that the differences in foreign and domestic market prices of risk are constant.

Table XI. Model 0 with Δ chosen by minimizing the CvM statistic.

Currency	α	β	b
Euro	-0.0912 (0.0217)	1.4244 (0.3060)	61
Yen	-0.1991 (0.0371)	0.4482 (0.0861)	35
Pound	0.0301 (0.0152)	-0.2407 (0.0600)	169
Swiss Franc	0.5209 (0.0429)	-0.7077 (0.1180)	39
Australian Dollar	0.0097 (0.0078)	-0.0216 (0.0351)	87
Canadian Dollar	0.0753 (0.0393)	-0.1070 (0.1155)	122

Table XII. Model 1 with Δ chosen by minimizing the CvM statistic.

Currency	α	β	γ	b
Euro	0.0185 (0.1392)	0.8999 (0.1330)	0.0000 (0.0001)	34
Yen	-0.1283 (0.1014)	1.1926 (0.1176)	0.0005 (0.0001)	123
Pound	0.0289 (0.0131)	1.6014 (0.0772)	0.0002 (0.0000)	180
Swiss Franc	0.1866 (0.0965)	2.0676 (0.2067)	-0.0009 (0.0001)	147
Australian Dollar	-0.1621 (0.0466)	1.0844 (0.1140)	0.0005 (0.0001)	89
Canadian Dollar	0.7395 (0.2244)	-0.0556 (0.4022)	-0.0004 (0.0001)	86

Table XIII. Model 2 with Δ chosen by minimizing the CvM statistic using the standard deviation of the full sample of the risky return as the measurement of risk.

Currency	α	β	γ	b
Euro	0.0017 (0.0162)	1.0032 (0.4995)	-0.0001 (0.0004)	63
Yen	0.0040 (0.0200)	0.4997 (0.2889)	-0.0002 (0.0009)	60
Pound	0.0031 (0.0122)	1.5695 (0.4559)	0.0001 (0.0006)	183
Swiss Franc	0.0864 (0.1068)	-0.2667 (0.3376)	-0.0006 (0.0011)	101
Australian Dollar	-0.0086 (0.0323)	1.0844 (0.2835)	0.0005 (0.0005)	42
Canadian Dollar	0.0067 (0.0057)	0.9331 (0.4172)	0.0000 (0.0001)	66

We next follow a similar approach to the slightly more realistic Model 2. Table XIII presents the results of our estimation for Model 2 using a variation of the Sharpe ratio. That is, we assume the actual return is the expected return and simply use the standard deviation of the risky return to be the volatility. The results are mixed. Only the estimate for the Swiss Franc is significantly different from unity. The yen and the Swiss Franc are not significantly different from zero and none are significantly negative. So, even in this extremely loose modeling for the market price of risk, we find no anomaly.

We next proceed to a more reasonable estimate of volatility. Our Table XIV

Table XIV. Model 2 with Δ chosen by minimizing the CvM statistic and with ν_t estimated using a rolling standard deviation for the previous month.

Currency	α	β	γ	b
Euro	-0.0002 (0.0074)	0.7937 (0.3085)	-0.0017 (0.0017)	34
Yen	-0.0097 (0.0034)	0.9907 (0.2662)	-0.0050 (0.0044)	54
Pound	0.0208 (0.0092)	0.9741 (0.2747)	-0.0023 (0.0020)	128
Swiss Franc	-0.1423 (0.1465)	0.5485 (0.3705)	0.0022 (0.0044)	147
Australian Dollar	0.0324 (0.0405)	0.8396 (0.5041)	-0.0061 (0.0086)	33
Canadian Dollar	0.2966 (0.0785)	1.0476 (0.0528)	-0.0006 (0.0004)	71

provides another estimate of Model 2 in which the volatility of the risky return is approximated using a rolling standard deviation for the previous month before any particular date. We see that none of the estimates are significantly different from one. Except for the Australian dollar and the Swiss Franc, every currency is significantly different from zero. Again, despite using this relatively simple model of risk, we see that the FPA disappears.

Finally, we consider Model 3, in which both market prices of risk may vary continuously, in Table XV. This model is perhaps the most realistic, and consequently, the estimates should be the most precise. In this case, we only model the market prices

Table XV. Model 3 with Δ selected by minimizing the CvM statistic and estimating risk via rolling standard deviation over the previous year.

Currency	α	β	b
Euro	-0.3401 (0.1241)	0.5753 (0.0961)	27
Yen	-0.0262 (0.0158)	0.8746 (0.0400)	145
Pound	-0.0498 (0.0511)	1.0479 (0.2048)	71

of risk for the euro, yen, and pound. We choose these as they are the most traded currencies and have the most developed domestic stock exchanges. All estimates are significantly different from zero. Only the yen is significantly different from one. Thus, in all three models and for all specifications, there is no evidence for the anomaly.

We have shown that, at least in the instantaneous horizon case, there is no evidence for the FPA. These results are in-line with the results of Chaboud and Wright (2005) and Bernoth, von Hagen, and de Vries (2007), that the forward premium anomaly only emerges as the maturity of the forward increases. We would expect this to be the case since $f_{t,t+\tau} - x_t \approx \int_t^{t+\tau} \pi_t ds$ for small τ . For short maturity forwards, the conventional FPA regression assumes that the market clears more frequently (i.e., daily rather than monthly). As the horizon becomes smaller, we would then expect the conventional FPA equation to be more appropriate and thus produce more accurate estimations.

4. Identification

The identification of these estimates can be assessed by examining the fragility of these estimates. Figure 9 shows the negative Cramér-von Mises distance for a range of possible values of the parameters β and γ while holding the constant term $\alpha = 0$ for the British pound in Model 1. This figure is typical for similar figures for all currencies. As can plainly be seen, the level of normality of returns, and therefore the degree of identification, is highly sensitive to the value of γ which corresponds to the constant term in the discrete time FPA equation. Conversely, the value of β has little impact on the CvM distance and is therefore poorly identified. On Figures 10, it is easier to discern behavior of the CvM distance as γ varies. The estimate which minimizes the CvM distance will depend critically on the estimate of γ . If the parameter is only slightly positive, the estimate of β will be positive, and vice versa. This is also typical of the estimates for all currencies, though the orientation is sometimes reversed.

5. Modeling the Market Price of Risk

Lastly, we will model the market price of risk in a more sophisticated way. There is currently no generally accepted method of accurately estimating the market price of risk. Moreover, developing an estimate of this latent variable is hardly a trivial task. With this in mind, we seek a balance between parsimonious estimates and more sophisticated estimates. Previously in this section, we have modeled the market price of risk using the Sharpe ratio with the standard deviation within a rolling window used as a proxy for volatility. Here, we venture into the more sophisticated techniques of estimation. Specifically, we nonparametrically estimate the market price of risk using standard kernel density estimation. This technique was demonstrated by Stanton

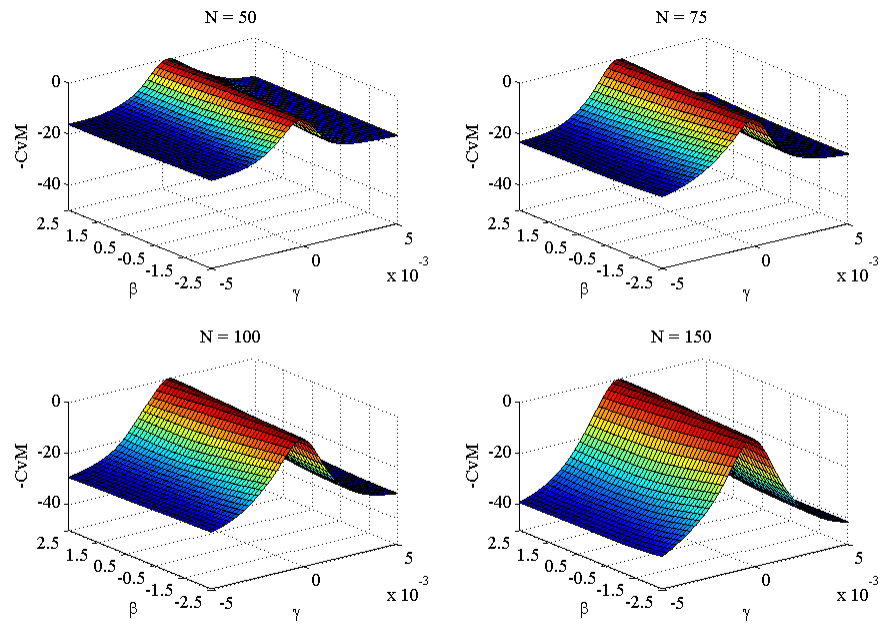


Fig. 9. The Cramér-von Mises distance for various values of the parameters β and γ holding $\alpha = 0$.

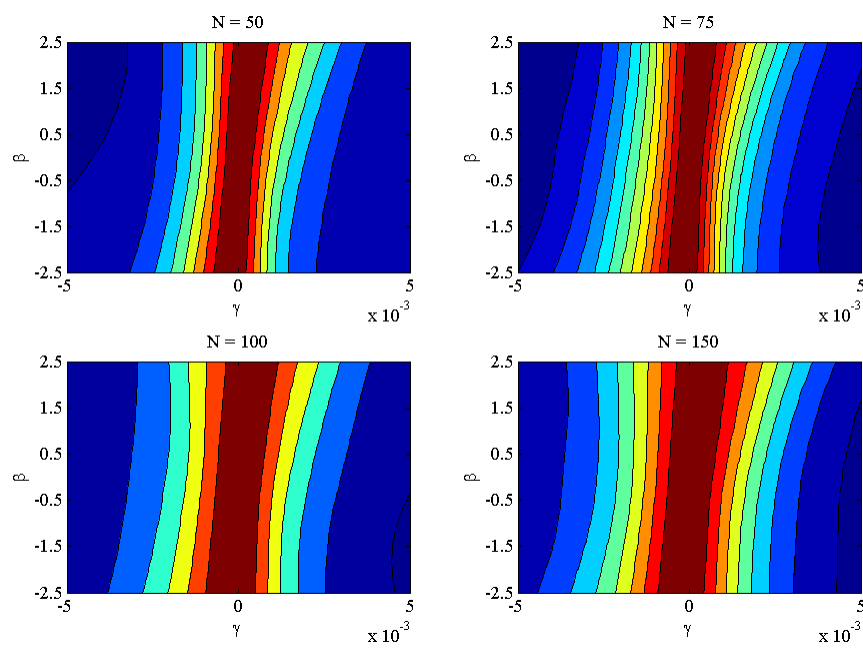


Fig. 10. A contour plot of the Cramér-von Mises distance for various values of the parameters β and γ holding $\alpha = 0$.

(1997).

Consider the exchange rate process

$$dX_t = \mu(X_t)dt + \sigma(X_t)dZ_t, \quad (3.24)$$

describing the behavior of the exchange rate process. As shown in Stanton (1997), we can estimate the drift and diffusion functions, up to the first order Taylor approximation, as

$$\mu(X_t) = \frac{1}{\Delta} \mathbb{E}_t [X_{t+\Delta} - X_t] + O(\Delta), \quad (3.25)$$

and

$$\sigma(X_t) = \sqrt{\frac{1}{\Delta} \mathbb{E}_t [(X_{t+\Delta} - X_t)^2]} + O(\Delta), \quad (3.26)$$

where Δ is the time between observations. If we further consider two exchange rate dependent assets, then the market price of risk, $\lambda(X_t)$ may be estimated as

$$\lambda(X_t) = \frac{\sigma(X_t)}{\Delta(\sigma^{(1)}(X_t) - \sigma^{(2)}(X_t))} \mathbb{E}_t [R_{t,t+\Delta}^{(1)} - R_{t,t+\Delta}^{(2)}] + O(\Delta), \quad (3.27)$$

where $\sigma^{(i)}$ is the diffusion term of asset i , and $R_{t,t+\Delta}^{(i)}$, $t + \Delta$ is the return on asset i in the period between t and $t + \Delta$.

Each of these conditional expectations may be readily approximated using simple kernel density estimation. For instance, to estimate $\mu(X_t)$, we may use

$$\mathbb{E}_t [X_{t+\Delta} - X_t | X_t = x] \approx \min_{\{a,b\}} \sum_{j=1}^n (Y_j - a - (X_j - x)b)^2 K\left(\frac{X_j - x}{h}\right), \quad (3.28)$$

where $K(z) = (2\pi)^{-1/2} e^{-(1/2)z^2}$ is the kernel function and h is the bandwidth. This is different from Stanton (1997) in that the local linear estimator is used rather than the local constant. We further depart from Stanton (1997) by setting the optimal bandwidth using the optimal bandwidth (see Bowman and Azzalini (1997), p.31). The two other assets we utilize are the US dollar invested at the annualized 3 month

Table XVI. Model 2 with Δ chosen by minimizing the CvM statistic and with ν_t estimated using the nonparametric technique from Stanton (1997).

Currency	α	β	γ
Euro	0.0175	1.4993	-0.0001
	(0.0743)	(0.3304)	(0.0001)
Yen	-0.0059	-0.1239	0.0001
	(0.0857)	(0.1905)	(0.0002)
Pound	0.0127	-0.4729	0
	(0.0381)	(0.015)	(0)
Swiss Franc	0.1097	-0.8662	0.0003
	(0.1025)	(0.9264)	(0.0002)
Australian Dollar	-0.2055	-0.2971	0.0001
	(0.0431)	(0.3159)	(0.0001)
Canadian Dollar	-0.2783	0.237	0.0002
	(0.1831)	(0.4571)	(0.0001)

treasury bill rate and the UK pound and invested at the annualized 3 month interbank rate. Note that from the domestic (US) point of view, this entails a spot transaction from dollars to pounds, investment of the pounds at the UK rate, and entering into a 12 month forward contract to convert the pound yield into dollars.

Table XVI gives the estimation results for Model 2 when the market price of risk is modeled in this way. The results seem to be more varying across currencies. Moreover, the troubles with identification are not improved, as can be seen in Figures 11 and 12.

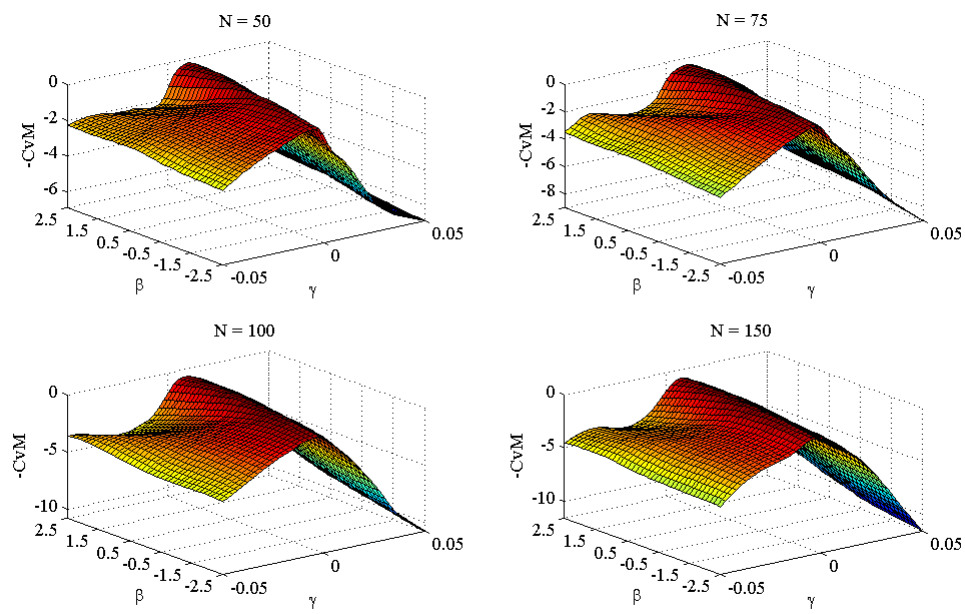


Fig. 11. The Cramér-von Mises distance for various values of the parameters β and γ holding $\alpha = 0$.

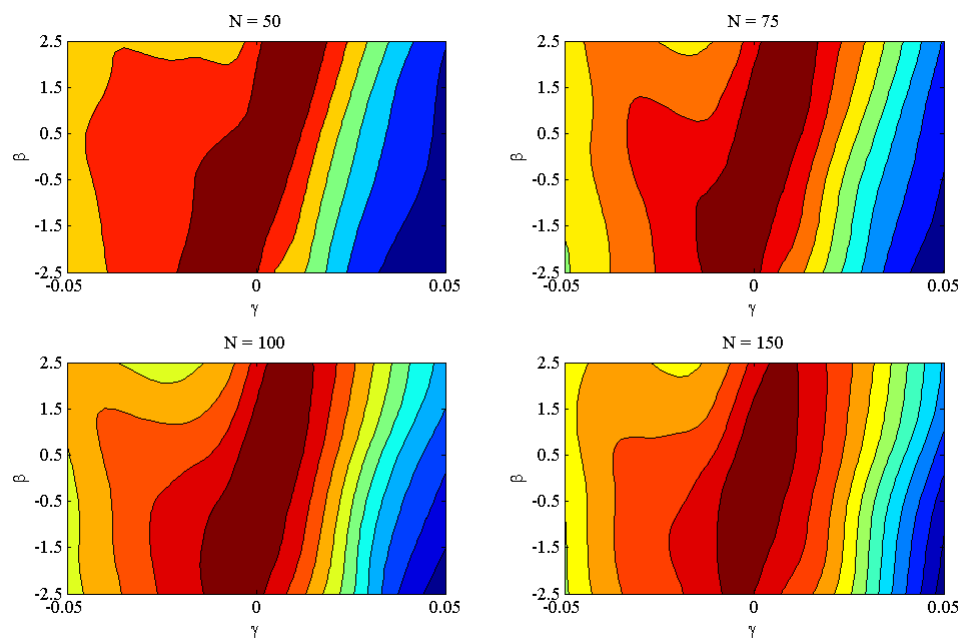


Fig. 12. A contour plot of the Cramér-von Mises distance for various values of the parameters β and γ holding $\alpha = 0$.

E. Conclusion

We have shown that, if the foreign exchange rate market clears continuously, the traditional FPA regression may be seriously misspecified. This would render traditional estimates invalid. To address this, we derive a new continuous time forward premium equation, linking the spot returns to both the instantaneous forward premium and the market prices of risk.

Foreign exchange rate data exhibit all of the complications commonly found in financial data. To correct for non-normality and stochastic volatility found in the exchange rate returns, we change from a standard clock to a volatility chronometer. Once this change is made, the return process should be identically and independently normally distributed. There exists the possibility that there remains endogeneity in our equation. We thus apply the Martingale Estimation Technique.

There remains several meaningful directions to advance this line of research. The next immediate expansion research is extending the analysis beyond the instantaneous forward exchange rate. This will be somewhat more economically complex since we can no longer use (3.14), and instead must derive a general continuous-time forward premium equation.

CHAPTER IV

CONCLUSION

As has been shown in the previous two chapters, observation in volatility time is an enormously powerful tool. Stochastic volatility is very detrimental to common statistical tests. Once the time change has been imposed, returns are independent, identically, and normally distributed. When combined with the Cauchy instrumental variable, all evidence for predictability disappears. Likewise, it becomes much easier to analyze the Forward Premium Anomaly. The lack of identification of the slope parameter becomes easily discernable.

REFERENCES

- AHN, S., S. FOTOPOULOS, AND L. HE (2001): “Unit Root Tests With Infinite Variance Errors,” *Econometric Reviews*, 20(4), 461–483.
- AÏT-SAHALIA, Y. (2004): “Disentangling Diffusion from Jumps,” *Journal of Financial Economics*, 74(3), 487 – 528.
- BACKUS, D. K., S. FORESI, AND C. I. TELMER (2001): “Affine Term Structure Models and the Forward Premium Anomaly,” *The Journal of Finance*, 56(1), 279–304.
- BAILLIE, R. T., AND T. BOLLERSLEV (1994): “The Long Memory of the Forward Premium,” *Journal of International Money and Finance*, 13(5), 565 – 571.
- BAILLIE, R. T., AND T. BOLLERSLEV (2000): “The Forward Premium Anomaly is Not as Bad as You Think,” *Journal of International Money and Finance*, 19(4), 471 – 488.
- BAILLIE, R. T., AND R. KILIÇ (2006): “Do Asymmetric and Nonlinear Adjustments Explain the Forward Premium Anomaly?,” *Journal of International Money and Finance*, 25(1), 22 – 47, Foreign Exchange Markets.
- BAKSHI, G., C. CAO, AND Z. CHEN (1997): “Empirical Performance of Alternative Option Pricing Models,” *Journal of Finance*, 52, 2003–2049.
- BALVERS, R., T. COSIMANO, AND B. McDONALD (1990): “Predicting Stock Returns in an Efficient Market,” *Journal of Finance*, 45, 1109–1128.

- BANK FOR INTERNATIONAL SETTLEMENTS (2007): *Triennial Central Bank Survey of Foreign Exchange and Derivatives Market Activity in 2007*. Basel: Bank for International Settlements Press & Communications.
- BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2006): “Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation,” *Journal of Financial Econometrics*, 4(1), 1–30.
- BEKAERT, G. (1996): “The Time Variation of Risk and Return in Foreign Exchange Markets: A General Equilibrium Perspective,” *Review of Financial Studies*, 9, 427–470.
- BEKAERT, G., AND R. J. HODRICK (1993): “On Biases in the Measurement of Foreign Exchange Risk Premiums,” *Journal of International Money and Finance*, 12, 115–138.
- BEKAERT, G., R. J. HODRICK, AND D. A. MARSHALL (1997): “On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates,” *Journal of Financial Economics*, 44, 309–348.
- BERNETH, K., J. VON HAGEN, AND C. G. DE VRIES (2007): “The Forward Premium Puzzle Only Emerges Gradually,” in *Tinbergen Institute Discussion Paper 07-033/2*.
- BJØNNES, G. H., AND D. RIME (2005): “Dealer Behavior and Trading Systems in Foreign Exchange Markets,” *Journal of Financial Economics*, 75, 571–605.
- BORODIN, A. N., AND I. A. IBRAGIMOV (1995): “Limit Theorems for Functionals of Random Walks,” *Proceedings of the Steklov Institute of Mathematics*, 195(2).

- BOSSAERTS, P., AND P. HILLION (1999): "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?," *Review of Financial Studies*, 12(2), 405–428.
- BOWMAN, A. W., AND A. AZZALINI (1997): *Applied Smoothing Techniques for Data Analysis*. Oxford, U.K.: Clarendon Press.
- BROWN, L., A. DASGUPTA, J. MARDEN, AND D. POLITIS (2004): "Characterizations, Sub and Resampling, and Goodness of Fit," *Lecture Notes-Monograph Series*, 45, 180–206.
- CAMPBELL, J. Y., AND J. H. COCHRANE (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205–251.
- CAMPBELL, J. Y., AND R. J. SHILLER (1991): "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *The Review of Economic Studies*, 58(3), 495–514.
- CAMPBELL, J. Y., AND M. YOGO (2006): "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81(1), 27 – 60.
- CAVALIERE, G. (2004): "Unit Root Tests Under Time-Varying Variances," *Econometric Reviews*, 23(3), 259–292.
- CAVALIERE, G., AND A. R. TAYLOR (2007): "Testing for Unit Roots in Time Series Models with Non-Stationary Volatility," *Journal of Econometrics*, 140(2), 919 – 947.
- CHABOUD, A. P., AND J. H. WRIGHT (2005): "Uncovered Interest Parity: It Works, but not for Long," *Journal of International Economics*, 66, 349–362.

- CHAN, Y., AND L. KOGAN (2002): “Catching Up with the Joneses: Heterogeneous Preferences and the Dynamics of Asset Prices,” *Journal of Political Economy*, 110(6), 1255–1285.
- CHANG, Y. (2002): “Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependency,” *Journal of Econometrics*, 110, 261–292.
- CHEN, W. W., AND R. S. DEO (2008): “Bias Reduction and Likelihood Based Almost-Exactly Sized Hypothesis Testing in Predictive Regressions Using the Restricted Likelihood,” Working Paper.
- CHUNG, H., AND J. Y. PARK (2007): “Nonstationary Nonlinear Heteroskedasticity in Regression,” *Journal of Econometrics*, 137(1), 230–259.
- COCHRANE, J. H. (2005): *Asset Pricing, (rev. ed.)*. Princeton, NJ: Princeton University Press.
- DAMBIS, K. E. (1965): “On the Decomposition of Continuous Semimartingales,” *Theory of Probability Applications*, 10, 401–410.
- DUBINS, L. E., AND G. SCHWARZ (1965): “On Continuous Martingales,” *Proceedings of the National Academy of Sciences*, 53, 913–916.
- DUFFIE, D. (2001): *Dynamic Asset Pricing Theory*. Princeton, NJ: Princeton University Press.
- ELLIOT, G., AND J. H. STOCK (1994): “Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown,” *Econometric Theory*, 10, 672–700.
- FOREIGN EXCHANGE COMMITTEE (2008): “Semi-Annual Foreign Exchange Volume Survey,” Discussion paper, New York Federal Reserve Bank.

- FROOT, K. A. (1990): "Short Rates and Expected Asset Returns," Working Paper 3247, National Bureau of Economic Research.
- GOODHART, C. A. E., P. C. MCMAHON, AND Y. L. NGAMA (1997): "Why Does the Spot-Forward Discount Fail to Predict Changes in Future Spot Rates?," *International Journal of Financial Economics*, 2, 121–129.
- GOYAL, A., AND I. WELCH (2003): "Predicting the Equity Premium with Dividend Ratios," *Management Science*, 49(5), 639–654.
- HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*. New York: Academic Press.
- HANSEN, L. P., AND R. J. HODRICK (1980): "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, 88(5), 829–853.
- HESTON, S. L. (1993): "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6(2), 327–343.
- HODRICK, R. J. (1989): "U.S. International Capital Flows: Perspectives from Rational Maximizing Models," *Carnegie-Rochester Conference Series on Public Policy*, 30, 231–288.
- HULL, J., AND A. WHITE (1987): "The Pricing of Options on Assets with Stochastic Volatilities," *The Journal of Finance*, 42(2), 281–300.
- JACEWITZ, S., H. KIM, AND J. Y. PARK (2008): "Testing for No Arbitrage in Continuous Time: A Resolution to the Forward Premium Anomaly," Working Paper.

- JACQUIER, E., N. G. POLSON, AND P. E. ROSSI (2004): “Bayesian analysis of stochastic volatility models with fat-tails and correlated errors,” *Journal of Econometrics*, 122(1), 185–212.
- KAMINSKY, G., AND R. PERUGA (1990): “Can a Time-Varying Risk Premium Explain Excess Returns in the Forward Market for Foreign Exchange?,” *Journal of International Economics*, 28(1-2), 47–70.
- KIM, T. H., S. J. LEYBOURNE, AND P. NEWBOLD (2004): “Behavior of Dickey-Fuller Unit-Root Tests under Trend Misspecification,” *Journal of Time Series Analysis*, 25(5), 755–764.
- LETTAU, M., AND S. LUDVIGSON (2001): “Consumption, Aggregate Wealth, and Expected Stock Returns,” *Journal of Finance*, 56(3), 815–849.
- LETTAU, M., AND S. VAN NIEUWERBURGH (2008): “Reconciling the Return Predictability Evidence,” *Rev. Financ. Stud.*, 21(4), 1607–1652.
- LONGWORTH, D. (1981): “Testing the Efficiency of the Canadian - U.S. Exchange Market Under the Assumption of No Risk Premium,” *Journal of Finance*, 36(1), 43–49.
- LORETAN, M., AND P. C. B. PHILLIPS (1994): “Testing the Covariance Stationarity of Heavy-tailed Time Series: An Overview of the Theory with Applications to Several Financial Datasets,” *Journal of Empirical Finance*, 1(2), 211–248.
- MARK, N. C., AND Y. MOH (2007): “Official Interventions and the Forward Premium Anomaly,” *Journal of Empirical Finance*, 14(4), 499–522.
- MENZLY, L., T. SANTOS, AND P. VERONESI (2004): “Understanding Predictability,” *Journal of Political Economy*, 112(1), 1–47.

- PARK, J. Y. (2002): “Nonlinear Nonstationary Heteroskedasticity,” *Journal of Econometrics*, 110, 383–415.
- (2008): “Martingale Regression and Time Change,” Working Paper.
- PARK, J. Y., AND P. C. B. PHILLIPS (1999): “Asymptotics for Nonlinear Transformations of Integrated Time Series,” *Econometric Theory*, 15, 269–298.
- PHILLIPS, P. C. B. (1987): “Towards a Unified Asymptotic Theory for Autoregression,” *Biometrika*, 74(3), 535–547.
- POLITIS, D., AND J. ROMANO (1993): “Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions,” *Annals of Statistics*, 89, 1303–1313.
- REVUS, D., AND M. YOR (2005): *Continuous Martingales and Brownian Motion (3rd ed.)*. New York: Springer-Verlag.
- SCHALLER, H., AND S. VAN NORDEN (1997): “Regime Switching in Stock Market Returns,” *Applied Financial Economics*, 7(2), 177–191.
- SO, B., AND D. SHIN (1999): “Cauchy Estimators for Autoregressive Processes with Applications to Unit Root Tests and Confidence Intervals,” *Econometric Theory*, 15, 165–176.
- STANTON, R. (1997): “A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk,” *Journal of Finance*, 52, 1973–2002.
- STÄRICĂ, C., AND C. GRANGER (2005): “Nonstationarities in Stock Returns,” *Review of Economics and Statistics*, 87(3), 503–522.

- TOROUS, W., R. VALKANOV, AND S. YAN (2004): “On Predicting Stock Returns with Nearly Integrated Explanatory Variables,” *Journal of Business*, 77(4), 937–966.
- VERONESI, P. (1999): “Stock Market Overreactions to Bad News in Good Times: A Rational Expectations Equilibrium Model,” *Review of Financial Studies*, 12(5), 975–1007.
- WELCH, I., AND A. GOYAL (2008): “A Comprehensive Look at The Empirical Performance of Equity Premium Prediction,” *Rev. Financ. Stud.*, 21(4), 1455–1508.
- WILCOX, S. E. (2007): “The Adjusted Earnings Yield,” *Financial Analysts Journal*, 63(5), 54–68.

APPENDIX A

MATHEMATICAL APPENDIX

Proof of Lemma 3.1 We let

$$\xi_i = \text{sgn}(x_{i-1})u_i,$$

and note that (ξ_i) is a martingale difference sequence such that

$$\begin{aligned} \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1}) &= \mathbb{E}(u_i^2 | \mathcal{F}_{i-1}) \\ \mathbb{E}\left(\xi_i^2 1_{\{|\xi_i| \geq \epsilon\sqrt{N}\}} \middle| \mathcal{F}_{i-1}\right) &= \mathbb{E}\left(y_i^2 1_{\{|u_i| \geq \epsilon\sqrt{N}\}} \middle| \mathcal{F}_{i-1}\right) \end{aligned}$$

Therefore, it follows from Assumption 3.1 that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \xi_i = \frac{1}{\sqrt{N}} \sum_{i=1}^N \text{sgn}(x_{i-1})u_i \rightarrow_d \mathbb{N}(0, \sigma^2) \quad (\text{A.1})$$

as $N \rightarrow \infty$, due to the standard central limit theorem (CLT) for martingale difference sequences in, e.g., Hall and Heyde (1980, Corollary 3.1, pp. 58-59). The proof for part (b) follows immediately, upon noticing that $y_i = u_i$ for all $i \geq 1$ if $\beta = 0$ under the null hypothesis of no predictability.

To establish part (a), we simply write

$$\kappa_N N^{-1/2}(\tilde{\beta}_N - \beta) = \left(\kappa_N^{-1} \sum_{i=1}^N |x_{i-1}| \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \text{sgn}(x_{i-1})u_i,$$

and note that we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \text{sgn}(x_{i-1})u_i = O_p(1)$$

from (A.1), and

$$\left(\kappa_N^{-1} \sum_{i=1}^N |x_{i-1}| \right)^{-1} = O_p(1)$$

from Assumption 3.2. The proof is therefore complete. \square

Proof of Theorem 4.2 Let $\beta = 0$, so that $y_i^* = u_i^*$ and $y_i^{*\delta} = u_i^{*\delta}$ for all $i \geq 1$. Then we define

$$\begin{aligned}\bar{\tau}(\tilde{\beta}_N^{*\delta}) &= \frac{1}{\sqrt{N\Delta}} \sum_{i=1}^N \text{sgn}(x_{i-1}^{*\delta}) u_i^{*\delta} \\ \bar{\tau}_0(\tilde{\beta}_N^{*\delta}) &= \frac{1}{\sqrt{N\Delta}} \sum_{i=1}^N \text{sgn}(x_{i-1}^{*\delta}) u_i^*.\end{aligned}$$

Since we assume

$$\hat{\sigma}_N^{*2} \rightarrow_p \sigma^{*2} = \Delta,$$

it suffices to show that

$$\bar{\tau}(\tilde{\beta}_N^{*\delta}) \rightarrow_d \mathbb{N}(0, 1) \tag{A.2}$$

as $N \rightarrow \infty$.

First, we show that

$$\bar{\tau}_0(\tilde{\beta}_N^{*\delta}) \rightarrow_d \mathbb{N}(0, 1) \tag{A.3}$$

as $N \rightarrow \infty$. To deduce (A.3), we let

$$\xi_i^* = \text{sgn}(x_{i-1}^{*\delta}) u_i^*.$$

Clearly, (ξ_i^*) is a martingale difference sequence. Moreover, it follows immediately that

$$\mathbb{E}(\xi_i^{*2} | \mathcal{F}_{i-1}) = \Delta$$

for all $i \geq 1$, and that

$$\mathbb{E}(|\xi_i^*|^k | \mathcal{F}_{i-1}) = \mathbb{E}(|u_i^*|^k) < \infty$$

for any $k \geq 0$, since (u_i^*) are iid $\mathbb{N}(0, \Delta)$. Consequently, we have as $N \rightarrow \infty$

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \xi_i = \frac{1}{\sqrt{N}} \text{sgn}(x_{i-1}^{*\delta}) u_i^* \rightarrow_d \mathbb{N}(0, \Delta),$$

due to the CLT for martingale difference sequences, and (A.3) follows immediately.

Second, we note that

$$\begin{aligned} \left| \bar{\tau}(\tilde{\beta}_N^{*\delta}) - \bar{\tau}_0(\tilde{\beta}_N^{*\delta}) \right| &\leq \frac{1}{\sqrt{N\Delta}} \sum_{i=1}^N \left| \text{sgn}(x_{i-1}^{*\delta}) u_i^{*\delta} - \text{sgn}(x_{i-1}^{*\delta}) u_i^* \right| \\ &= \frac{1}{\sqrt{N\Delta}} \sum_{i=1}^N |u_i^{*\delta} - u_i^*| = o_p(1), \end{aligned}$$

due to the results in Lemma 4.1. Therefore, we have

$$\bar{\tau}(\tilde{\beta}_N^{*\delta}) = \bar{\tau}_0(\tilde{\beta}_N^{*\delta}) + o_p(1),$$

from which and (A.3) we have (A.2). The proof is therefore complete. \square

Proof of Proposition 4.3 Write

$$\begin{aligned} \tau(\tilde{\beta}_N^{*\delta}) &= \frac{1}{\hat{\sigma}_N^* \sqrt{N}} \sum_{i=1}^N \text{sgn}(x_{i-1}^{*\delta}) y_i^{*\delta} \\ &= \frac{\beta}{\hat{\sigma}_N^* \sqrt{N}} \sum_{i=1}^N |x_{i-1}^{*\delta}| + \frac{1}{\hat{\sigma}_N^* \sqrt{N}} \sum_{i=1}^N \text{sgn}(x_{i-1}^{*\delta}) u_i^{*\delta}. \end{aligned} \quad (\text{A.4})$$

Note that $(\text{sgn}(x_{i-1}^{*\delta}) u_i^{*\delta})$ is a martingale difference sequence and $(\text{sgn}(x_{i-1}^{*\delta}) u_i^{*\delta})^2 = u_i^{*\delta^2}$ for all $i \geq 1$. Consequently, we have

$$\mathbb{E} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \text{sgn}(x_{i-1}^{*\delta}) u_i^{*\delta} \right)^2 = \frac{1}{N} \sum_{i=1}^N \mathbb{E}(u_i^{*\delta^2}) < \infty,$$

due to condition (a) of Assumption 4.3, and it follows that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \text{sgn}(x_{i-1}^{*\delta}) u_i^{*\delta} = O_p(1). \quad (\text{A.5})$$

The stated result may now be easily deduced from (A.4) and (A.5), due to condition (b) of Assumption 4.3. This completes the proof. \square

APPENDIX B

ADDITIONAL PREDICTABILITY RESULTS

Table XVII shows the effects of a structural break in volatility on the standard OLS t -test and the time-changed Cauchy t -test. The negative impact on the standard t -ratio may easily be seen. For even a single break in volatility, the effects are even more dramatic than for the previous stochastic volatility models. For $\sigma_1 = 3$ and $\sigma_2 = 1$, nearly 70% of the time, the standard t -ratio will reject a true null at the 10% significance level and nearly a quarter at the 1% level. Again, the time-changed Cauchy t -ratio exhibits none of these problems. The actual size is very close to the nominal size in all cases. Moreover, the distribution is symmetric as the left-, right-, and two-tailed tests all have similar rejection rates. It is also worth noting that the size of the break in no way affects the size of the test.

Table XVIII presents the results of those experiments. The distortionary effects may easily be seen. Under the Heston model, for the simple two sided test, at the 1% nominal level, the null is rejected more than 4% of the time. Our simulation results for the Heston model are quite similar the Cavaliere and Taylor (2007) with 29% rejection at the 5% nominal level. The distortion is even more apparent when we examine the left and right tail tests. In fact, the standard OLS predictability test will reject a true null at the 10% level almost half of the time. Likewise, the distortion is just as severe using exponential stochastic volatility, with as much as 48% being rejected at the 10% nominal level and as much as 9% being rejected at the 1% level.

Table XVII. This table compares results based on the standard OLS regression to the results from Cauchy estimation with a time change when there is a structural break in volatility. Here, “TC-Cauchy” represents the Cauchy estimate of the time-changed data and $T = 25$. All values are given in terms of percentages.

	$\sigma_1 = 1.5$ and $\sigma_2 = 1$		$\sigma_1 = 3$ and $\sigma_2 = 1$	
	OLS	TC-Cauchy	OLS	TC-Cauchy
Two Sided Test				
10%	41.04	10.20	51.46	10.83
5%	27.37	5.14	36.79	5.43
1%	9.86	1.32	15.68	1.22
Left Tail Test				
10%	0.02	9.69	0.00	10.48
5%	0.00	4.83	0.00	5.31
1%	0.00	1.11	0.00	1.17
Right Tail Test				
10%	58.64	10.46	68.42	10.42
5%	41.04	5.37	51.46	5.52
1%	15.43	1.26	23.50	1.23

Table XVIII. This table compares results based on the standard OLS regression to the results from Cauchy estimation with a time change. Here, “TC-Cauchy” represents the Cauchy estimate of the time-changed data and $T = 25$. All values are given in terms of percentages.

	Heston Model		Exponential SV	
	OLS	TC-Cauchy	OLS	TC-Cauchy
Two Sided Test				
10%	29.38	10.22	31.11	9.77
5%	17.11	5.53	18.96	4.85
1%	4.44	1.06	5.53	0.84
Left Tail Test				
10%	0.46	9.78	0.37	9.81
5%	0.07	5.00	0.10	4.84
1%	0.01	0.99	0.01	0.95
Right Tail Test				
10%	46.50	10.60	48.21	10.05
5%	29.31	5.22	31.01	5.05
1%	8.27	1.25	9.55	0.87

APPENDIX C

ADDITIONAL FPA ESTIMATION RESULTS

Table XIX reports the post time-change number of observations N which yields the lowest, and therefore most normal, one dimensional CvM statistic for each model and currency. It also reports the median estimate of β across all levels of N . Note that the selected N 's are not close to the minimum boundary of 30. Thus, it is reasonable to conclude that the constraint is not binding. Also notice that all of the median estimates are positive, other than Model 0. Moreover, after Model 0, they are all close to unity.

Table XIX. This table provides the selected values of N and equivalently Δ for each country based on minimizing the one dimensional CvM statistics of parameter estimates. The median estimates of β across the range of Δ is also reported.

Currency	Model 0		Model 1	
	N	Median $\hat{\beta}$	N	Median $\hat{\beta}$
Euro	78	1.4988	45	1.5154
Yen	58	0.4440	135	0.86144
Pound	189	-0.3951	199	1.1486
Swiss Franc	50	0.3939	177	1.1241
Australian Dollar	101	-0.0166	101	0.93395
Canadian Dollar	179	-0.57-6	116	0.72029
Currency	Model 2		Model 3	
	N	Median $\hat{\beta}$	N	Median $\hat{\beta}$
Euro	73	1.3612	37	1.5598
Yen	79	0.5628	155	0.6394
Pound	195	0.9499	81	0.5323
Swiss Franc	198	0.9733	—	—
Australian Dollar	61	0.6569	—	—
Canadian Dollar	84	0.7075	—	—

APPENDIX D

THE NELSON-SIEGEL PROCEDURE

The instantaneous forward premium is obviously not observable. We estimate π_t using the well know Nelson-Siegel procedure. Forward exchange rate premiums as a function of maturity have shapes which are invariably monotonic, humped, or ‘S’ shaped. If we assume spot exchange rates are generated by a differential equation, then forward exchange rates, being forecasts, will be the solution to the equations. Thus, the forward exchange rate at time t , f_t , as a function of the time until maturity τ , would be of the form:

$$f_t(\tau) = \beta_{1t} + \beta_{2t}e^{-\lambda_t\tau} + \beta_{3t}\lambda_te^{-\lambda_t\tau}$$

where β_{it} are parameters to be estimated and the term λ_t represents the decay rate. Such a model readily produces the shapes required of forward exchange rate curves. This is simply the functional form of a forward interest rate applied to forward exchange rates. The Nelson-Siegel procedure is a widely used parsimonious method for interpolating forward interest rates. The Nelson-Siegel approximation is especially popular among central banks for estimating forward interest rates.

We follow Diebold and Li (2006) and choose approximately 30 months as a medium term maturity. We can then use the corresponding constant value of λ_t which they give as $\lambda_t = 0.0609$. This allows the model to be estimated using ordinary least squares. This simplifies and adds numerical robustness since many potentially challenging numerical optimizations are replaced with trivial least squares. Using Nelson-Siegel approximation, we can estimate a β for every possible maturity for each day t .

VITA

Stefan A. Jacewitz majored in mathematics and economics at the University of Oklahoma and received his B.A. degree in August 2004. Later that month, he entered the doctoral program in economics at Texas A&M University. He served as a National Science Foundation Fellow in the GK-12 program for the Fall of 2005 and Spring of 2006. He was awarded teaching and research assistantships from the Fall of 2005 until Spring of 2009. The research of this dissertation was granted the Incentives for Excellence fellowship in 2008. In August 2009, Dr. Jacewitz was awarded a Ph.D. specializing in financial econometrics. He is currently employed as a financial economist by the Center for Financial Research at the Federal Deposit Insurance Corporation in Washington, D.C. He may be reached via the Economics Department at Texas A&M University at 4228 TAMU, College Station, TX 77843. His current email is jace1neo.tamu.edu.

The typist for this thesis was Stefan A. Jacewitz.